

Understanding Addiction

by

Robert M. Kelly

April 22, 2021

A dissertation submitted to the
faculty of the Graduate School of
the University at Buffalo, The State University of New York
in partial fulfillment of the requirement for the
degree of

Doctor of Philosophy
Department of Philosophy

Copyright by
Robert M. Kelly
2021
All rights reserved

To my dad, who never stopped pushing me to be better.
To my mom, who never stopped believing that I could be.
To Jo, Chet, Lily, and Bob.
And to Mar-Mar. You are still missed every day.

Acknowledgements

Many people contributed to this project in one way or another. First, thank you to my dissertation supervisor, Professor David Hershenov, for the incredible support and guidance you provided throughout my graduate career. Your encouragement and enthusiasm about my work helped me push through numerous difficult bumps in the road. Many thanks also to my two other committee members, Professors Barry Smith and Neil Williams. This project would be nothing without ontology and dispositions, and I could not have asked for two better people to advise me on these topics. The three of you have helped me more than I can begin to articulate here, both with regard to this project and well beyond it. I have been both honored and fortunate to have had you three in my corner.

I also owe an enormous amount of gratitude to the fellows and friends of the Romanell Center for Clinical Ethics and Philosophy of Medicine (including those who were around when it was still PANTC). The feedback and the philosophical engagement with my work at many conferences, workshops, and dinners was absolutely instrumental in helping to refine the ideas in this project. For their helpful comments on papers that became chapters and for discussions over core ideas within the dissertation, special thanks to Harvey Berman, Jim Delaney, Neil Feit, David Hershenov, John Keller, Steve Kershner, Rasmus Larsen, David Limbaugh, Jake Monaghan, Phil Reed, and Jerry Wakefield.

Thank you to Dr. Susanne Uusitalo for your many insights and the inspiration you provided. You may not know it, but our many wonderful conversations about addiction during the Free Will summer school in Latvia were the initial catalyst for my starting this project.

I owe thanks to Professors James Beebe and Claire White for always encouraging me to do interdisciplinary work and for helping me to hone that skill. It was undeniably influential in this project.

Thanks also to Dr. Paul Meyer for allowing me to invade your seminar and learn more about addiction in one semester than I thought possible. Thanks also for allowing and encouraging me to do so much philosophy of addiction in a psych seminar on experimental models of mental disorders.

Thank you to the University at Buffalo Research Institute on Addictions (RIA), and especially Linda Hughes, DNP, Molly Rath, LCSW, and Dr. Kim Walitzer. My time volunteering at the RIA and working with you three was an invaluable practical experience for a mere philosopher struggling to engage with and the clinical literature and to understand practical approaches to studying addiction.

I am also grateful to the faculty and graduate body of the Department of Philosophy at the University at Buffalo for helping to create an engaging environment in which I was able to pursue, express, and develop the ideas that became this project, and to the staff for making whatever administrative hurdles that arose seem easy to clear. Of my current and former fellow graduate students, I am greatly indebted, in particular, to Brendan Cline, Francesco Franda, Danielle Limbaugh, David Limbaugh, Eric Merrell, Jake Monaghan, and Ariane Nomikos. I am truly a better philosopher and a better person than I would otherwise have been, were it not for your support, your insights, and your friendships. Thank you for helping me make it through to this milestone.

Finally, thank you to all of my family and friends for the unending support through this long journey. I know I would not have been able to accomplish this feat without that support. In particular, some special thanks are in order. First and foremost, thank you to my parents, Matt and Sue Kelly. You made this possible and I am forever grateful. Thank you also to Brandy Kelly (Carrizales). The love, support, and sacrifice you gave as I pursued this dream are forever in my heart. Thanks to Julian Barton and Andrew Rios – no one has known such loyal and supportive friends. Thanks to Bill Keene for the uncountable conversations that I subjected you to about my work and philosophy in general. They were more helpful than you can know. Lastly, a world of thanks to Megan Cross for your unrelenting love, understanding, and encouragement, without which I would not have survived the final stretch of this project over the past year and a half. I am so happy, and sometimes still in awe, that the universe saw fit to bring our paths side by side once again.

Abstract

The addiction literature is fraught with conceptual confusions, stalled debates, and an unfortunate lack of clear and careful attempts to delineate the phenomenon of addiction in a way that might lead to consensus. My dissertation has two overarching aims, one metaphysical and one practical.

The first aim is to defend an account of addiction as the *systematic disposition to fail to control one's desires to engage in certain types of behaviors*. I defend the inclusion of desires and impaired control in the definition, and I flesh out the notion of *systematicity* central to my dispositionalist framework. I engage the so-called 'disease vs. choice' debate, criticizing its presupposition that we are dealing here with a dichotomy and arguing that the movement towards a middle ground is the right track to take. I explain how the dispositionalist account can capture this middle ground and how it serves to expand upon existing views, in particular by filling in the metaphysical details.

The second aim is to show how the account I defend can help to unify the extant views and disciplinary perspectives in the literature. Both the dispositionalist aspect of my framework and the methodology adopted (applied ontology and systematic metaphysics) can move the literature towards both substantive and methodological unification. This will help to clear up conceptual confusions, resolve (or sometimes *dissolve*) apparently intractable disputes, situate different research perspectives with respect to each other, facilitate interdisciplinary dialogue, and help to frame important questions about addiction. Finally, I offer the beginnings of an ontology of addiction, which will provide a terminologically well-structured guide to the addiction literature in a way that will facilitate more effective and efficient communication and data management across disciplines.

Table of Contents

Acknowledgements	iv
Abstract	vi
List of Tables	xii
List of Figures	xiii
Chapter 1: Towards a General Account of Addiction	1
1.1 Introduction and Project Aims	1
1.1.1 Goal One: Providing an Account of Addiction	2
1.1.1.1 Downfalls of Lacking a Satisfactory Account	2
1.1.1.2 How Addiction Works vs. What Addiction Is	3
1.1.2 Goal Two: Helping to Unify Addiction Research.....	5
1.1.2.1 Unification Through Dispositionalism.....	6
1.1.2.2 Unification through Ontology.....	8
1.2 Assumptions and Terms	9
1.2.1 Distinguishing Ontology from Ontologies	10
1.2.2 A Brief Introduction to Dispositions	12
1.2.3 Addiction as a Disposition	13
1.2.3.1 Terms Related to ‘Addiction’	14
1.3 The Need for a General Account of Addiction	15
1.3.1 Disunification in the Addiction Literature	17
1.3.1.1 Vague or Unclear Definitions of Terms.....	18
1.3.1.2 Potayto, Potahto: Term Usage Differs between Authors	22
1.3.1.3 Some Definitions Are of Low Quality.....	25
1.3.1.4 Interpreting the Evidence for Disunification	27
1.3.1.5 Two Consequences: Shoddy Work Rolls Downhill.....	28
1.3.1.6 A Quick Recap	30
1.3.2 Ontology as the Solution to the Disunification Problem	31
1.3.2.1 Realist Ontologies and the Semantic Web	31
1.3.2.2 Controlled Vocabularies Facilitate Interdisciplinary Research	33
1.3.2.3 What a General Account Comes To	34
1.3.3 Desiderata for a General Account of Addiction.....	35
1.3.3.1 Empirical Sensitivity.....	36
1.3.3.2 Theoretical Breadth	38
1.3.3.3 Proper Level of Analysis.....	39
1.3.3.4 How these Desiderata Capture those from West and Colleagues.....	39
1.3.4 A Dispositionalist Account as the Foundation for Unification	41
1.4 Lessons and Looking Ahead	43
Chapter 2: Ontology and Unification in Addiction Research	46
2.1 Introduction and Chapter Road Map	46
2.1.1 The Arguments	47
2.2 Ontology Counteracts Disunification	49
2.2.1 A Contextual Primer: The Need for Ontology Is Ubiquitous and Unsurprising	49
2.2.1.1 The Need for Intervention	50

2.2.1.2 Contextual Features of the Need for Ontology: Science and Heterogeneity	52
2.2.2 The First Goal of Ontology: Carving Up the World at Its Joints	53
2.2.2.1 Entities and Relations	54
2.2.2.2 Continuants, Occurrents, and the Sub-Type Relation	56
2.2.2.3 An Upper-Level Ontology and More Relations	58
2.2.2.4 Taxonomies: The Backbone of Ontologies	61
2.2.2.5 After the <i>is_a</i> : Aristotelian Definitions	62
2.2.2.6 Universals and Defined Classes	64
2.2.2.7 Whales Are Not Fish: Providing a Shared Vocabulary	65
2.2.2.8 Goal One and Unification	67
2.2.3 The Second Goal of Ontology: Organizing Our Findings.....	69
2.2.3.1 Heterogeneous Data Calls for Ontology	69
2.2.3.2 The First Type of Organization: Classification.....	72
2.2.3.3 The Second Type of Organization: The Semantic Web.....	73
2.2.4 Third Goal of Ontology: Representing the World.....	77
2.2.4.1 Knowledge Is Not Reality: Realism and Conceptualism.....	77
2.2.4.2 Let's Get Real: Realism Is a Better Standard Practically	78
2.2.4.4 Additional Principles of Best Practice	82
2.2.5 What the Goals of Ontology Have Taught Us.....	83
2.3 Unification through Ontology: Two Success Stories.....	84
2.3.1 Just GO with It: Ontology for Biology	84
2.3.1.1 The Glue of the GO: Some Unifying Features	85
2.3.1.2 GO-ing Towards Unification.....	88
2.3.2 Two Thumbs for OGMS: Ontology for General Medicine	88
2.3.2.1 Some Unifying Features of OGMS	89
2.3.2.2 OGMS and Unification	92
2.4 Applying Ontology to Addiction Research	92
2.4.1 What's in a Discipline? The No Difference Argument	93
2.4.2 Prototype Addiction Ontologies	95
2.5 Lessons from Chapter 2.....	95
<i>Chapter 3: Towards a Dispositionalist Account of Addiction.....</i>	<i>98</i>
3.1 Introduction and Chapter Road Map.....	98
3.2 The Limitations of Ontology's Unifying Power.....	99
3.2.1 Two Senses of 'Unification': Methodological and Substantive	99
3.2.2 Accepting AAA Would Lock Ontology Out of Substantive Unification	103
3.2.3 If Not Ontology, then What?	105
3.3 The Basics of Being Disposed	106
3.3.1 Disposition as a Familiar Notion	106
3.3.1.1 Ordinary Cases of Dispositions	107
3.3.1.2 Modal vs. Categorical Properties.....	108
3.3.1.3 Two Things Dispositions Tell Us.....	109
3.3.2 Defining 'Disposition'	110
3.3.3 Speaking of Dispositions: 'Powers', 'Capabilities', and Other Dispositionalist Terms	113
3.3.3.1 Two Points of Clarification.....	113
3.3.3.2 Powers: The Most Fundamental of Dispositions	114
3.3.3.3 Capabilities and Functions: Dispositions of Interest	116
3.3.3.4 Predispositions: Dispositions to Dispositions	119
3.3.4 Dispositions <i>for</i> Something: Triggering Conditions, Realizations, and Manifestations	121

3.3.4.1 Multi-Track Dispositions	122
3.3.4.2 Direct vs. Indirect Manifestations and the Standard of Reality	123
3.3.5 Dispositions are Internally Grounded in their Bearer’s Physical Makeup	124
3.3.6 Three Senses of Reliability: Strength, Opportunity, and Systematicity	125
3.3.6.1 Reliability as Strength: The Likelihood of Being Realized when Triggered	126
3.3.6.2 Reliability as Opportunity: The Commonality of Triggering Conditions	126
3.3.6.3 Reliability as Systematicity: Sufficiently Strong in the Right Triggering Conditions.....	127
3.3.7 Taking Stock of the Dispositional Features.....	130
3.4 The Addiction Disposition: Desires, Impaired Control, and Systematicity	131
3.4.1 Desires as a Component of Addiction	133
3.4.1.1 What Desires Are	134
3.4.1.1.1 Reward Learning and Desires	135
3.4.1.1.2 Unconscious Desires, Habits, and Appetites.....	139
3.4.1.2 The Role of Desires in Addiction.....	141
3.4.2 Impaired Control as a Component of Addiction.....	143
3.4.2.1 Initial Motivation and the Misleading Disease vs. Choice Debate.....	143
3.4.2.2 Further Reason to Require Impaired Control	145
3.4.3 Addiction as a Systematic Loss of Control.....	148
3.4.3.1 How We Can Fail to Control Desires	148
3.4.3.2 Systematicity and Societally Relevant Circumstances	151
3.4.3.3 Types of Control: A Theory-Neutral Framework.....	155
3.5 A Recap of Chapter 3’s Arguments	158
<i>Chapter 4: The Power of the Dispositionalist Account of Addiction</i>	<i>160</i>
4.1 Introduction and Chapter Road Map.....	160
4.2 Capturing Key Phenomena with the Dispositionalist Account of Addiction.....	162
4.2.1 The Puzzle of Addiction	163
4.2.1.1 The Brain Disease Model’s Explanation of the Puzzle	165
4.2.1.2 Pickard’s Explanation of the Puzzle	167
4.2.2 Addiction Is Janus-Faced	169
4.2.3 Addiction Is Context-Sensitive.....	171
4.2.4 Addiction Is Sometimes Indeterminate.....	173
4.2.5 Addiction Takes Many Forms	174
4.2.6 ‘Addiction’ Is Polysemous.....	176
4.2.7 Addiction as It Can Be.....	178
4.2.7.1 Harmless Addictions	179
4.2.7.2 Instant Addicts	179
4.2.7.3 Addictions that Never Show Themselves	180
4.2.7.4 Alien Addictions	182
4.2.8 The Dispositionalist Account Is Explanatorily Powerful	183
4.3 The Dispositionalist Account of Addiction Captures Extant Views	183
4.3.1 A Disposition by Any Other Name	184
4.3.1.1 Someone Had to Say It	184
4.3.1.2 Some Accounts Are More Explicitly Dispositionalist	186
4.3.1.3 Most Accounts Are Implicitly Dispositionalist	188
4.3.2 Capturing A Potluck of Views	192
4.3.3 A Verbal Disagreement Exposed	195
4.3.4 A Unifying Account	197
4.4 Ontologizing the Dispositionalist Account.....	198

4.4.1 Adhering to Ontology Best Practices	199
4.4.1.1 Proper Definitional Form	199
4.4.1.2 A Consistent “is a” Hierarchy	201
4.4.1.3 Realism and BFO	202
4.4.2 Representing the Components of the Dispositional Account in a BFO-Compliant Way	203
4.4.2.1 Addiction as a Disposition in BFO	204
4.4.2.2 Addicts Are Organisms Bearing the Addiction Disposition	204
4.4.2.3 Addiction Is Realized in Uncontrolled Behaviors	205
4.4.2.4 Attempts to Control One’s Desires in Societally Relevant Choice Situations	207
4.4.3 The Ontological Usefulness of the Dispositionalist Account	209
4.5 Defending the Bridge Premise (3e) of the Argument for Premise (3)	211
4.6 Lessons and Looking Ahead	213
Chapter 5: How the Dispositionalist Account Fares Against Extant Accounts.....	215
5.1 Introduction and Chapter Road Map.....	215
5.2 Evaluation Criteria	217
5.2.1 Substantive Desiderata: How to Conceptualize Addiction	217
5.2.2 Methodological Desiderata: How to Represent Addiction in an Ontology	222
5.3 Evaluating Extant Accounts of Addiction	226
5.3.1 George Ainslie	228
5.3.2 Bennet Foddy and Julian Savulescu.....	230
5.3.3 Nick Heather.....	232
5.3.4 Gene Heyman.....	234
5.3.5 Neil Levy	235
5.3.6 Marc Lewis.....	238
5.3.7 Hanna Pickard.....	240
5.3.8 Nora Volkow	242
5.3.9 Jerome Wakefield.....	244
5.3.10 Summary: Extant Accounts are Impaired or Metaphysically Insufficient.....	246
5.4 Evaluating Definitions in Existing Classification Systems and Vocabularies.....	247
5.4.1 Ontological Representations of Addiction.....	249
5.4.1.1 CLO and SNOMED-CT: No Definition Provided	250
5.4.1.1.1 The Classification in CLO	251
5.4.1.1.2 The Classification in SNOMED-CT	252
5.4.1.2 CHIRO, MP, ONTONEO, and uPHENO: A Dependence with Certain Results	254
5.4.1.3 DOID: Mental Disorder Involving Substance Abuse or Dependence	256
5.4.1.4 EFO: State of Dependence or Disorders Related to Drug Exposure	258
5.4.1.5 MFOMD: A Harmful Mental Disease (or Disorder? or Disease Course?).....	260
5.4.1.6 NBO: Appropriately Caused Behavioral Response to Addictive Substance	261
5.4.2 Non-Ontology Classification Systems and Vocabularies on ‘Addiction’	263
5.4.2.1 The DSM-V: A Cluster of Criteria	264
5.4.2.2 The ICD-11: Classifying Addiction	267
5.4.2.3 Medical Vocabularies: The NCI, NIDA, and the DHHS	271
5.4.3 Extant Resources Representing Addiction are Flawed	274
5.4.4 AddictO to the Rescue?	276
5.5 Where We Stand: Summarizing the Arguments of the Foregoing Chapters	279
References.....	283

Appendix A: Hyperbolic Discounting and Addiction.....305
Appendix B: OBO Foundry Principles of Best Practice for Ontology307
Appendix C: BFO Principles of Best Practice for Ontology.....309

List of Tables

<i>Table 1: Number of Cases of Coronavirus Disease at Hospitals</i>	<i>71</i>
<i>Table 2: Substantive Desiderata (Account of Addiction)</i>	<i>218</i>
<i>Table 3: Methodological Desiderata (Ontology)</i>	<i>223</i>
<i>Table 4: Violations of Substantive Criteria by Extant View</i>	<i>247</i>
<i>Table 5: Summary of DSM-V Criteria for Alcohol Use Disorder</i>	<i>266</i>
<i>Table 6: Violations of Substantive and Methodological Criteria by Resource</i>	<i>275</i>

List of Figures

<i>Figure 1: Entity Types in BFO</i>	58
<i>Figure 2: Representation of the is_a and part_of Relations</i>	59
<i>Figure 3: Example Diagram Representing Multiple Types of Entities and Relations</i>	60
<i>Figure 4: Example of a Complex Ontology Diagram</i>	60
<i>Figure 5: The Five Kingdom Classification System of Living Organisms</i>	61
<i>Figure 6: Example of a Possible Ontology Classifying Whales as Fish</i>	66
<i>Figure 7: DISEASE and DISEASE COURSE are Distinct Classes in OGMS</i>	90
<i>Figure 8: Relations between DISEASE, DISORDER, and Related Entity Types</i>	90
<i>Figure 9: Components of a Disease Course</i>	91
<i>Figure 10: Illustration of (Part of) Reward-Based Learning</i>	136
<i>Figure 11: Illustration of (Part of) the Reward Calculation Process</i>	138
<i>Figure 12: Addiction as a Disposition in BFO</i>	204
<i>Figure 13: Addict as the Bearer of the Addiction Disposition</i>	205
<i>Figure 14: Addiction as Realized in Addicted Behavior</i>	206
<i>Figure 15: Addiction as Triggered by Attempts to Control Desires in Choice Situations</i>	208
<i>Figure 16: Societally Relevant Choice Situations as Related to the Addict's Reference Class</i>	209
<i>Figure 17: No Definition for the SNOMED CT Class Addiction</i>	250
<i>Figure 18: Meaning of Definition Status for Addiction in SNOMED CT</i>	251
<i>Figure 19: Classification of Addiction in CLO</i>	251
<i>Figure 20: Classification of Addiction in SNOMED-CT</i>	253
<i>Figure 21: Classification of Addiction in MP</i>	255
<i>Figure 22: Classification of Addiction in ONTONEO</i>	256
<i>Figure 23: Classification of Drug Dependence in EFO</i>	259
<i>Figure 24: Classification of Addiction Disorder in MFOMD</i>	261
<i>Figure 25: Classification of Addiction Response in NBO</i>	262
<i>Figure 26: Classification of Alcohol Dependence in ICD-11</i>	269
<i>Figure 27: Classification of 'addiction' in NCIt</i>	272
<i>Figure 28: Top-Level (Root) Terms in AddictO</i>	278

Chapter 1: Towards a General Account of Addiction

1.1 Introduction and Project Aims

This is a dissertation about addiction. In it I attempt to accomplish two overarching goals. First, I defend my own account of what addiction is. I argue for the view that addiction is, broadly speaking, a certain type of disposition. Thus, I defend a *dispositional account of addiction*. I think many addiction researchers do this as well, even though they are commonly either not explicit about or not aware that they are doing so. Thus, one novelty that my account brings to the table is the adoption of a *strictly* dispositionalist framework for talking (hopefully unambiguously) about addiction.¹

Second, I aim to provide an account (and approach) that will serve as a genuinely useful addition to the vast and ever-growing literature on addiction. I argue that both the dispositionalist aspect of the framework itself, as well as part of the methodology adopted (in particular, incorporating the methods of ontology; more on this below and in the next chapter), will help to clear up confusions in the literature, resolve (or sometimes *dissolve*) apparently intractable disputes, better situate different research perspectives with respect to each other, and facilitate interdisciplinary dialogue.

In this opening chapter, I accomplish four tasks. First, I say a bit more about each of my two overarching goals of the project. They are closely related and understanding them will help to bring the different parts of the project into a unified whole. Second, I introduce and define the central terms and assumptions. Third, I argue that a general account of addiction is what the field of addiction

¹ Note that ‘strictly’ here is meant to demarcate an account like mine from those that are less explicit about understanding addiction as a disposition, as well as less detailed about the metaphysics. The contrast class, then, consists in those who are either closet dispositionalists about addiction, whether intentionally or not, or simply unclear about what sort of thing they take addiction to be. An example of a non-dispositionalist approach would be one that counts addiction as a type of behavior or process (perhaps a cyclical process). Another might be to hold that addiction is a social construct. It may seem surprising that I am treating the claim ‘addiction is a disposition’ as something new. I expound on why my claim is accurate, even if surprising, in later chapters (especially Chapter 4).

research needs, and I introduce three important desiderata for a successful account of addiction. Fourth, I summarize the core lessons of the chapter and offer a quick look at the chapters that follow.

1.1.1 Goal One: Providing an Account of Addiction

As is well known, addiction is one of those phenomena that, on its surface, can seem relatively clear to us, but is in fact quite puzzling. For instance, we take ourselves to know that addiction exists – it seems that there are *obviously* addicts in the world.² This is generally a starting assumption (one which I adopt), and the ease with which we take it on board can lead to feeling like we also know pretty clearly *what* addiction is, *who* the addicts are, and so forth. Despite this façade of lucidity, addiction can seem incredibly hard to pin down. It does not take long for the attempt to move from *addicts existing in the world* to then *clearly and correctly identifying them* to become exceedingly difficult. Succeeding in the latter task, though, is important.

1.1.1.1 Downfalls of Lacking a Satisfactory Account

For instance, one serious problem that can arise concerns the empirical data we have on addicts and addiction, whether genetic, neurobiological, psychological, behavioral, sociological, or otherwise. In order to obtain empirical data studies must be conducted, and in order to conduct studies the researchers must construct or utilize inclusion criteria for experimental and control groups. In other words, they must decide who does and who does not count as an addict. Of course, this presupposes that we already know what addiction is, and thus how to tell who the addicts (and non-

² There are what might be called ‘addiction skeptics’ who speak and write about the so-called ‘myths of addiction’ (Davies, 1997; Dill & Holton, 2014; Foddy & Savulescu, 2010a, 2010b; Ley, 2014). However, it is almost always more accurate to describe such authors as ‘addiction-as-disease skeptics’, or perhaps ‘addiction-as-*brain*-disease skeptics’. They interpret the brain disease view (the most dominant view, accepted virtually across the board in clinical contexts) as entailing that addicts are slaves to their desires in some unique way that fully undermines their agency. These skeptics are typically only expressing a distaste for this definition of ‘addiction’ and commonly hold that, while some people are addicts, there is no difference in kind between addiction and ordinary forms of reward-driven, *akratic* behavior. Even those who see addiction as simply a strong form of *akrasia* (Heather, 2017a, 2020) have criticized the fully skeptical *myth of addiction* view (Heather & Segal, 2015). I discuss *no difference* views more in Chapters 4 and 5.

addicts) are. What else would it mean for researchers to put one group of subjects into the *addiction* group and others into the *control* or *non-addict* group?

As it happens, a huge amount of existing empirical data utilizes (quite unabashedly) the criteria offered in the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (DSM-V).³ As I explain below, this creates its own problem because the DSM-V criteria do provide a satisfactory account of what addiction is.⁴ The point is that without a careful and satisfactory account of addiction, we risk running into a self-fulfilling prophecy concerning what we know about addicts and the nature of addiction. As Nick Heather has elegantly pointed out, the addiction literature is well below-average with respect to the clear and careful use of definitions, even definitions of ‘addiction’:

When reading literature on addiction, both in the popular media and in scientific publications, it is surprising how seldom authors actually tell us what they mean by the term, whether by a formal definition, a rough characterization of how they see addiction, or what they consider to be the “hallmark” (Skog 1999, p. 173) of addictive behavior and experience. It is as though authors simply assume that, when speaking of addiction, everybody will know what they mean. This might have been understandable in the early days of addiction studies in the 1950s or 1960s, when the word conjured up images in most people’s minds of a disheveled individual injecting heroin into a vein or desperately seeking the means to do so. Most people would agree that such stereotypes should have been abandoned long ago but there nevertheless remains a tendency for writers on addiction to assume, for whatever reason, that the term needs no clarification.⁵

1.1.1.2 How Addiction Works vs. What Addiction Is

But addiction is also puzzling in another way. Addicts are very often ambivalent about their addicted behavior.⁶ They will continue to engage in such behavior not only despite significant costs, but despite having (and even despite having vocalized) the genuine desires, intentions, and resolutions to cease doing so. This has been called the ‘puzzle of addiction’, and it is often the target (or catalyst) for various theories of addiction and addicted behavior.⁷ Indeed, some authors seem to think that

³ I discuss these criteria in detail in Chapter 5. See Sanislow et al. (2020) for discussion of using DSM criteria for research.

⁴ Summers (2015) argues for a similar point.

⁵ Heather (2017b, p. 3).

⁶ Many authors discuss this point (Foddy & Savulescu, 2010a; Heyman, 2009, 2013; Hodgson et al., 2020, Ch. 7). That addicts are ambivalent is also a fundamental assumption behind the use of *motivational interviewing* in treating addiction (Feldstein Ewing et al., 2016; Miller & Rollnick, 2013).

⁷ Notable examples are Heather (2017b) and Pickard (2016, 2019, 2020).

what it means to give an account of addiction is just to explain the puzzle of addiction.⁸ This is what appears to call out for explanation, after all. Why would someone continue engaging in behavior that is harmful and that they explicitly claim they want to stop doing?

But while the puzzle of addiction is important and fascinating in its own right, even if solving it can provide us with an account of addiction (which I think is questionable), it still cannot provide us with a *comprehensive* account. For instance, consider that not every case of addiction involves ambivalence, as is seen in the case of willing addicts.⁹ Another problem this puzzle raises is that, in attempting to solve it, some accounts confuse things like *evidence of addiction* or *how addiction works* with addiction itself. For instance, neither particular behavioral patterns nor characteristic neurobiological changes are themselves the addiction. They might be causes or effects of an addiction (or perhaps symptoms, if addiction is a disease or disorder), but explaining how addiction works and how addicts behave is not the same thing as providing an account of what addiction is.¹⁰

The point of the first goal, then, is to clearly and carefully delineating what addiction is, starting out from the fundamental assumption that some people are addicts.¹¹ I take myself to know that this

⁸ For instance, proponents of the brain disease model of addiction (BDMA) typically take themselves to be solving this puzzle in providing their account (Pickard, 2019). Many examples of the BDMA exist (R. Z. Goldstein & Volkow, 2002; Hyman, 2005; Kalivas & Volkow, 2005; G. F. Koob et al., 2014; G. F. Koob & Le Moal, 2006; A. Leshner, 1999; A. Leshner & Koob, 1999; Alan I. Leshner, 1997; N. Volkow, 2014; N. Volkow et al., 2016; N. Volkow & Fowler, 2000; N. Volkow & Koob, 2015; N. Volkow & Morales, 2015; Wise, 2000). See Kincaid & Sullivan (2010) for a review of BDMA accounts.

⁹ Willing addicts have been widely discussed in the addiction literature (Fischer & Ravizza, 2000; Flanagan, 2013, 2016; Foddy & Savulescu, 2010a; Frankfurt, 1971; Kennett, 2013; Sinnott-Armstrong & Pickard, 2013; Sripada, 2017; Sripada & Railton, 2018). Kennett is somewhat skeptical of willing addicts, but there is clinical evidence that willing and resigned addicts exist (Pickard & Pearce, 2013). However, Flanagan (2013) argues that shame is essential to addiction, suggesting ambivalence is necessary for addiction. Though, he later provides examples of willing addicts, implying that he still thinks they exist, and shame is simply *compatible* with their existence (Flanagan, 2016). In any event, willing addicts are possible, even if none presently exist, which is all I need for the point to land. I come back to willing addicts in Chapter 5.

¹⁰ This is a substantive claim, and I revisit it briefly below and then again in more detail in Chapters 4 and 5.

¹¹ A skeptical reader might ask what it would take, on my view, for this assumption to be falsified. Is it possible to discover that there are no addicts (as opposed to discovering that many people are wrong about what addiction is)? Perhaps. Though, this might miss the point. I am interested in providing an account of a particular phenomenon that I take to be the proper referent of the term 'addiction'. I think this phenomenon should be and often is (albeit implicitly) what we refer to when we speak of addiction. Moreover, failing to recognize this has been problematic for the literature, as I go on to explain in later chapters. The bottom line is that there *is* some difference-making phenomenon that 'addiction' gets at. Jim Carrol *really was* different, in some way, from Mother Theresa with respect to using heroin (or pick your favorite examples of a clear addict and non-addict). Addiction researchers are trying to say what this difference comes to. With this in mind,

is true, and I assume that you do, too. Of course, sometimes we might be wrong about who the addicts are. Nonetheless, whatever turns out to be true regarding who the addicts are, what they are addicted to, what their brains are doing, and so forth, I know that there are addicts. However, clearly not everyone is an addict. So, my most basic assumption becomes: There are addicts and there are non-addicts. This, it seems, is obvious.¹²

So, with this assumption in mind, the central question guiding this project is: What makes the difference? That is, what is it that makes an addict an addict, and not a non-addict? This is not so obvious. A heroin addict's relationship to their addicted behavior differs in an important way from their relationship to playing basketball or to drinking water. Similarly, that relationship is importantly different from the *non-addict's* relationship to using heroin. An account of addiction should tell us what makes this difference. Many accounts attempt to provide us with such an answer – to tell us about the nature of addiction. However, many end up answering a different question, such as: How does addiction work? It is the former question which I aim to answer in the chapters that follow.

1.1.2 Goal Two: Helping to Unify Addiction Research

It is important to say why my account is a useful addition to the literature, especially given the abundance of theories of addiction that already exist. The main reason can be stated rather simply: Despite the plethora of accounts, the literature needs unifying, and my account offers two ways to help do that. Above, I mentioned other benefits such as helping to clear up confusions, dissolve disputes, situate research perspectives, and so forth. What these really come to, though, is *unification* in

the question seems to ask whether it was possible that Jim Carrol and Mother Theresa really *were not* different, in any way, with respect to using heroin. I doubt it, but I also doubt that this question is very useful once we see the project of addiction researchers in this way.

¹² Some might object that everyone is an addict in *some* sense since we are all addicted to something or other, whether coffee or exercising, or perhaps eating food and drinking water (Dill & Holton, 2014). I agree that it is *possible* that everyone is an addict. Still, an addict is always addicted to *something*, and so I can amend my claim to there being addicts and non-addicts with respect to some type of behavior. Some people are heroin addicts, and some are not. Thus, there would be a meaningful difference between addicts and non-addicts even if everyone were addicted to a specific behavior.

different forms. I discuss unification in more detail below. Here I only want to briefly clarify the main idea behind this second overarching aim.

It is sometimes claimed in philosophical work that everyone else has the wrong view about some issue p , and the proposed thesis is going to set us straight with the right view about p : “This person claims that p is a , b , and c , while that person claims that it is x , y , and z ...but I am here to tell you that it is actually $l-m-n-o...p$.” Of course, there is often overlap in views, and competitors are said to have come close or to have gotten some pieces of their account right. A common result is that virtually everyone else’s view ends up on either end of some spectrum, and the correct view is said to lie somewhere in the middle, partially made up of lifted parts as if coming directly out of a chop shop. The account defended here can be seen as pushing towards the middle in a similar way, and I will also argue that the account can capture much of what the prominent extant accounts get right (in addition to avoiding many of their mistakes).

However, what is more interesting about my account, and what allows for a better claim to unification, is that I think most accounts of addiction are largely right in a fundamental sense. You sometimes have to squint to see it, but it seems to me that most authors are basically talking about the same thing. They are often either describing things differently or (unknowingly) describing a different aspect of the same thing, which is confusing for everyone and liable to mislead us into thinking that disagreement about the nature of addiction abounds. Interestingly, though, authors on addiction are often not disagreeing. At least, that will be my contention, and my second aim is to show that my account helps to make this clear – it unifies existing models of addiction, and it does so in two ways.

1.1.2.1 Unification Through Dispositionalism

First, the fact that my account of addiction is dispositionalist will help to capture most other theories and approaches. In this sense, it is *substantively unifying*. The reason it can do this can, again, be

put quite simply: Most accounts are dispositionalist, even though most authors are unaware of this. Sometimes they focus on different parts or aspects of the addiction disposition (such as its underlying neurobiology or behavioral manifestations). This is a bit like what is suggested in the old metaphor of the blind men and the elephant.¹³ Each blind man studies (that is, feels) a different part of the elephant (though he does not know what he is feeling) and is asked to give an account of his object of study. Each comes up with a different account of what an elephant is: a tree for the man feeling the leg; a fan for the man feeling an ear; a snake for the man feeling the trunk; and so on. Of course, they are in fact all studying different aspects of the same thing. However, the point I want to make is also saying much more than this.

A common lesson from the elephant metaphor has a relativistic bent: Addiction is a trunk, and it is also a tail, and a leg, and an ear, since it just depends on how you look at it.¹⁴ My claim is that the story is more like an analogy that suggests an objective *convergence* of views than a metaphor that suggests a relativistic *equalizing* of views.¹⁵ Addiction is not brain changes, and behavioral patterns, and persistent desires, and harmful habits, and hyperbolic discounting all at the same time. Instead, addiction is a disposition (or so I will argue), where brain changes, behavioral patterns, desires, habituation, harm, discounting, and the like all have their place within the dispositionalist framework. Some things might partially constitute the disposition, while others end up being its manifestations (such as behavioral patterns), its triggering conditions (such as environmental conditions), or its causal story (such as repeated use or pre-existing conditions and tendencies). My aim is to provide and explain the framework, including the vocabulary, that will help to reveal that which underlies these phenomena – namely, *addiction*, the different aspects of which most extant accounts seem to be getting at.

¹³ For a translation of one of the oldest appearances of the metaphor, see Bhikkhu (2012). The metaphor is widely used in the addiction literature. For some paradigmatic examples, see Du Plessis (2014a) and Peele (1987).

¹⁴ For a more relativistic interpretation of the metaphor, see Henderson (2000, p. 3) and Kalant (1994).

¹⁵ For what it is worth, the latter (relativistic equalizing) also naturally facilitates disciplinary information siloing, while the former does not. As I go on to explain, this counteracts unification within and between research disciplines.

1.1.2.2 Unification through Ontology

Second, my methodology includes the application of the methods and guiding principles of what is called ‘applied ontology’ (or, as in its typical non-philosophical context, and henceforth, just ‘ontology’). Implementing ontology will facilitate the integration of the varied research perspectives on addiction. In this sense, my account is also *methodologically unifying*. Unlike the men in the elephant metaphor, addiction researchers often have some access to and interaction with each other – more specifically, with each other’s theories, terms, data, methods, and so on, which is of course typically welcomed. Good researchers want their work to be informed by other disciplines, and in particular by the best science and theory from the various relevant disciplines. The same is true in addiction research. However, the central difficulty this raises is that there is a need for interdisciplinary conversation while there is yet no settled interdisciplinary language. Researchers must talk to each other, but at present they cannot do so – at least not very well.

Ontology is designed to help solve this problem. The use of the methods and principles of ontology to solve this very same problem in the domains of biology and medicine is well-known.¹⁶ It involves the use of controlled vocabularies with which researchers across disciplines can communicate (theorize, converse) and annotate or tag their data. This facilitates more effective and efficient communication and data management. This is why a handful of addiction researchers who have done work in or are familiar with ontology have recently both called for ontology to be more widely used to help clean up the literature,¹⁷ and also spearheaded a series in the journal *Addiction* (called ‘Addiction

¹⁶ See discussions of the success of the Gene Ontology, Ontology for General Medical Science, and ontology generally (Arp et al., 2015; Ashburner et al., 2000; Bada et al., 2004; Ceusters & Smith, 2010, 2015; Courtot, 2016; The Gene Ontology Consortium, 2019; S. E. Lewis, 2017; Scheuermann et al., 2009; B. Smith, 1995; B. Smith & Ceusters, 2010, 2015).

¹⁷ Some have called for ontology work on addiction specifically (Cox et al., 2020; du Plessis, 2018; Hastings, le Novere, et al., 2012a; West, Christmas, et al., 2019a; West et al., n.d.; West, Marsden, et al., 2019a) while some have called for ontology in closely related fields like human behavior interventions and mental health more generally (Hastings & Schulz, 2012; Larsen et al., 2017; Michie et al., 2017; Michie & Abraham, 2008; Michie & Johnston, 2017; Wright et al., 2020).

Theories and Constructs’) that is meant to provide a shared forum for addiction researchers to help realize the aim of developing addiction ontologies.¹⁸

The point of the second goal, then, is securing substantive and methodological unification as so described. The former is secured through the dispositionalist account of the nature of addiction defended here. I will argue that this account not only captures what is true about addiction, but that it also captures what most other accounts are getting at. The latter is achieved through the inclusion of ontology into the methodological approach adopted here. I will argue that the aforementioned authors are right to call for ontology building in addiction research, and that the account I defend constitutes a significant step towards adhering to the methods and principles of good ontology building.

1.2 Assumptions and Terms

In this section I introduce some of the central terms that will be used, as well as some of the fundamental assumptions that the account rests on. Five preliminary points are in order. First, some of the definitions here will be refined as the account is fleshed out and I attempt to more thoroughly elucidate their intended referents. Second, the definitions here are therefore mostly meant to get us started off on the same page. Third, I focus only on the most central concepts here and provide definitions for other terms as needed when we come to them. Fourth, while this is not a glossary, it might be a useful reference if a question arises as to what I mean when I use any of these terms. Given the above quote from Nick Heather about the problem with definitions in addiction research, this section is meant to put that worry to bed. Lastly, here are some typographical conventions that will be used in this dissertation:

- (i) Section headings are numbered according to chapter, section, sub-section, sub-subsection, and so on from left to right; thus, ‘1.1.1.1’ refers to Chapter 1, section 1, sub-section 1, sub-sub-section 1, and ‘2.2.2’ refers to Chapter 2, section 2, sub-section 2, and so on down the line;

¹⁸ See West et al. (2019b).

- (ii) I use **COPPERPLATE** font to refer to types, universals, or classes in discussing ontology;
- (iii) I use *italics_with_underscores* to refer to types of relations in discussing ontology;
- (iv) The use of *italics without underscores* will indicate an emphasis (sometimes because the term is important and being used for the first time, sometimes to emphasize the definition or meaning of a term, and sometimes simply for literary emphasis);¹⁹ and,
- (v) I will use single quotes to indicate I am mentioning a term, such as ‘disposition’.
- (vi) I will use the following number schemes for lists and arguments:
 - a) English numerals are used for premises in my main argument (such as (1), (2), ...), and letters are added for sub-premises supporting them (such as (1a), (1b), ...);
 - b) English numerals following ‘P’ and ‘C’ are used when reconstructing other authors’ arguments (such as (P1), (P2), (C1), ...);
 - c) Small Roman numerals are used for lists (such as (i), (ii), ...), and lower-case English letters are used for sub-point in lists (such as a), b), ...)

1.2.1 Distinguishing Ontology from Ontologies

I use the term ‘ontology’ in two ways. First *ontologies* are controlled vocabularies meant to represent the types and structure of entities and relations in some given domain. In this sense, they are “representational artifacts.”²⁰ The vocabulary combines easy-to-understand language with formal, computational definitions that computers can reason over. They start with a backbone taxonomy that sets out the hierarchy of entities in terms of type-subtype relations. For instance, the type **HUMAN BEING** is a *subtype_of* (or ‘*is_a*’) **MAMMAL**, which *is_a* **ANIMAL**, which *is_a* **MATERIAL ENTITY**, and so on. Next, they include certain other relations and properties that the entities stand in or bear. Expressions representing entities and relations, and their corresponding definitions, are added, amended, or sometimes removed as the ontology expands or is modified to cover (represent) more portions of reality, or to do so more accurately or in more detail. In this sense, an ontology is a living

¹⁹ An exception might be when I use the title of a document, such as when I discussed the journal *Addiction* above. To keep emphases distinct from reference to types of relations, I will always use an underscore to separate terms for the latter, and therefore, will always use two-or-more-word labels for relation types (such as ‘*is_bearer_of*’ instead of ‘*bears*’).

²⁰ Ceusters and Smith (2010, p. 4).

entity subject to change as research develops.²¹ Ontologies can be domain-specific, for instance representing only types of *biological* entities and relations. Ontologies can also be domain-neutral, representing the most general types of entities, properties, and relations that span many domains of knowledge (such as **MATERIAL ENTITY, QUALITY, PROCESS**, *has_part, is_bearer_of*). Here are some descriptions of ontologies from the ontologists themselves:

Ontologies, from this perspective, are representational artifacts, comprising a taxonomy as their central backbone, whose representational units are intended to designate *universals* (such as *human being* and *patient role*) or *classes defined in terms of universals* (such as *patient*, a class encompassing *human beings* in which there inheres a *patient role*) and certain relations between them.²²

...within the Semantic Web units of information are explicitly defined in shared vocabularies or “ontologies” – computational representations of knowledge in a particular domain. Ontologies link together the words, expressions, and language that humans use to refer to things, with computable formal definitions of those things that allow computers to distinguish one type of thing from another.²³

Second, *ontology* refers to the practice of building ontologies for the purpose of providing clarity, logical coherence, and other means for facilitating interdisciplinary dialogue and research. This is the sense in which some people are said to *do* ontology (which has ontologies as its output), and it is what I will do in part of this dissertation. Unfortunately, not all ontologists are doing ontology in exactly the same way.²⁴ For that reason, I should note that I follow the Open Biological and Biomedical Ontologies (OBO) Foundry’s realist approach to ontology building, the methods and guiding principles of which are founded in the upper-level Basic Formal Ontology (BFO).²⁵ The most important idea here is the *realism* being endorsed. This means that ontologies are assumed to represent portions of reality (as opposed to our concepts or knowledge of reality). As humans are fallible,

²¹ For instance, see Buttigieg et al. (2013). The previous sentence also assumed that ontologies are realist – that is, its terms are intended to represent entities and relations in reality. I will defend this claim in the next chapter.

²² Ceusters and Smith (2010, p. 4).

²³ West, Christmas, et al. (2019a, p. 164).

²⁴ This is unfortunate because lack of unity in ontology building, such as excessive “multiplication of ontologies,” will only create some of the same problems of clarity and unification found in the literature that ontology is meant to solve in the first place (Arp et al., 2015, pp. xvi–xvii).

²⁵ These principles are outlined on the OBO Foundry’s homepage and are described in more detail by Barry Smith, original developer of BFO, and his colleagues (Arp et al., 2015; The OBO Foundry, 2020b). They appear in Appendices B and C.

sometimes the ontologies they build will fail to accurately represent the relevant domain of reality.²⁶ Nonetheless, the assumption is that what goes into the ontology is believed with good scientific evidence to represent some portion of reality.

1.2.2 A Brief Introduction to Dispositions

Dispositions are modal properties.²⁷ A *property* is a way something is, and *modality* is concerned with possibility and necessity. To somewhat oversimplify, dispositions are *the ways an entity bearing them is that explain what that entity can or must do in certain circumstances*.²⁸ Consider some simple examples.

A vase is the kind of thing that will break when dropped on a hard surface, at least under normal conditions. It is *disposed* to break when it is suitably struck (we know this disposition by the name ‘fragility’). Salt dissolves when it is put into water. It is *disposed* to dissolve under those conditions (we know this disposition by the name ‘solubility’). Why is salt water-soluble and why are vases fragile? In short, because of their dispositions. Lady Gaga can play the piano. This is true even while she is on a plane from New York City to Los Angeles. Were there a piano on the plane, Lady Gage could play a song if she tried. Why? Again, because of her dispositions. Tiger Woods is an expert putter. He is *disposed* to make attempted putts within a reasonable distance from the hole, under normal conditions. He can still miss a gimme putt from time to time, but his expertise is what explains our thinking

²⁶ See Ceusters & Smith (2010, pp. 3–4) for discussion of referring and non-referring units of an ontology.

²⁷ I draw heavily on Williams (2019) in thinking about dispositions. While Williams is a *powers theorist*, which is to say that he holds a particular view about the nature and fundamentality of some dispositions, one need not be a powers theorist to endorse a dispositionalist account of addiction. All that is needed is the presence of dispositions in some sense in one’s ontology or metaphysic. This is a minimal requirement. Many versions of powers theories about dispositions exist (Aristotle, 2001; Heil, 2012; Jacobs, 2017; E. J. Lowe, 2006; Marmodoro, 2010; McKittrick, 2018; Molnar & Mumford, 2006; Mumford & Anjum, 2011; Reid, 1788; Vetter, 2015; Williams, 2019). For the neo-Humean view that dispositions are explained by true counterfactual conditionals, see Lewis (2001). Chapter 3 provides more details about dispositions.

²⁸ Dispositions might never manifest, such as when we securely bubble wrap a vase and lock it away never to be touched again. This does not destroy the vase’s *fragility* (a disposition). Its fragility is precisely why we wrap and lock it up.

(correctly) that he *could have* (even *should have*, in the predictive sense) made it.²⁹ Why? You know the tune by now – because of his dispositions.

So, dispositions are *that which explain what an entity bearing them can or must do in certain circumstances, under normal conditions*.³⁰ As we can see in the above examples, there are three main components to dispositions. First, there is the disposition itself (fragility, solubility, piano playing expertise, putting expertise), which is the *way* the entity is such that it is disposed towards doing such and such. Second, there is the *manifestation* of the disposition, which is *what the entity can or must do* in virtue of having that disposition (breaking, dissolving, playing piano, sinking putts). These are the processes it *can* produce or bring about. Third, there are the *manifestation* (or ‘triggering’) *conditions*, which are *those conditions which, together with the disposition, help produce the manifestation* (being suitably struck, being submerged in water, attempting to play the piano, attempting to sink a putt).

1.2.3 Addiction as a Disposition

Defining ‘addiction’ will comprise much of the dissertation, but in the spirit of clarity and directness, let me spoil the ending some. Addiction is a certain type of disposition. This means, at least, that an addiction will have certain types of manifestations and manifestations conditions, just as do *fragility*, *solubility*, and the rest. Moreover, addiction will have a certain type of profile, so that when a person bears the disposition *addiction*, she will be some way. For *fragility* and *solubility*, we might suppose, their bearers will have a certain molecular structure. But the *piano playing* or *putting expertise*

²⁹ Tiger Woods has certainly missed a short putt, but due to some relatively abnormal (and frustrating) circumstances: <https://www.youtube.com/watch?v=KYAQT6fTZAc>. See also Austin (1956) for a famous discussion of putting and ability to do otherwise, and Dennett (2004) for a supportive follow up on Austin’s example.

³⁰ Of course, entities might be disposed towards certain states or processes were they to be in *abnormal* conditions. The distinction here is more about the difference between a vase’s being fragile because, roughly, it will break when dropped, and a bowling ball’s *not* being fragile even though it will break when dropped from the Empire State Building. Also, I say ‘can produce’ because being disposed need not guarantee manifestation, even in normal conditions. A disposition may be probabilistic, where the manifestation conditions only produce some probability of the disposition manifesting. Others may be deterministic, manifesting with certainty provided conditions are normal. A disposition may also be present but unable to manifest in ordinary conditions, as with a bubble-wrapped vase locked away in a vault.

enjoyed by humans seem quite different. These are more complex dispositions that seem to involve having other underlying dispositions like *hand-eye coordination*, *muscle control*, *balance*, and so forth. In this sense, addiction is much more like being a pianist or an expert putter than being fragile or soluble. It is a complex disposition, with a suite of underlying dispositions that make it up.

But we are getting ahead of ourselves. Addiction is the disposition to undergo uncontrolled choice behaviors – intending to drink, deciding to buy cigarettes, choosing to forego tonight’s *Narcotics Anonymous* meeting, and so on – with some sufficient likelihood in a sufficient number of certain types of manifestation or triggering conditions.

1.2.3.1 Terms Related to ‘Addiction’

Here are some further definitions relating to the understanding of addiction. An *addict* is an entity that bears the addiction disposition. Typically, these are people. Possibly, addicts are sometimes animals like mice and rats.

Addicted behavior is the type of behavior that is characteristic of an addict, behavior in which he engages in virtue of bearing the addiction disposition.³¹ Some examples are *failing to resist an urge to drink*, *continuing to gamble despite harmful consequences*, and *smoking a cigarette despite a recent genuine resolution to quit*. In addicts, such behaviors will usually, but not necessarily, be manifestations of the addiction disposition.

An *object of addiction* is whatever an addict is addicted to, be it a type of substance or type of behavior.³² Addicts are always addicted to something. For instance, an addict might be disposed to

³¹ I use ‘addicted behavior’ rather than ‘addictive behavior’ because the latter implies that the behavior itself causes addictions by its nature. The former (I hope) more clearly suggests that the relevant behavior is only *that which is typical of addiction*. The type of behavior will correspond with the type of addiction. Heroin addicts will have behavior types like *seeking out heroin*, *using heroin*, and *deciding to use heroin* as their addicted behaviors. Sugar addicts’ addicted behaviors will pertain to seeking and consuming sugar, and so on.

³² Compare this with the phrase ‘addiction object’ used by West et al. (2019a), which is slightly more inclusive. What’s more, ‘object of addiction’ is just a helpful way of speaking. On my view, addicts are addicted to types of behaviors, not objects/substances. This is because, as we will see in later chapters, it is processes (such as behaviors) that constitute the manifestations of a disposition (such as addiction). Still, it is perfectly sensible to say things like ‘Jim is addicted to heroin.’”

addicted behaviors concerning heroin and its consumption, or gambling, or exercise. These would be the object of addiction in each case, and they make sense of the phrases ‘heroin addict’, ‘gambling addict’, and ‘exercise addict’.³³

An *addiction component* is any part of the addiction disposition.³⁴ This includes any underlying properties (dispositional or otherwise) that make up or constitute the addiction disposition, as well as manifestations, manifestation conditions, or parts thereof.

1.3 The Need for a General Account of Addiction

In this section, I motivate the need for developing a general account of addiction. I also lay out three desiderata for successfully providing such an account and argue that ontology is the best route to satisfying them.³⁵ In recent work, Robert West and colleagues identify a research challenge within the study of addiction involving the need for more clarity and unity within the addiction literature.³⁶ We can call this the ‘addiction research challenge’. Moreover, they argue that solving the addiction research challenge is best done through the use of the Semantic Web, the backdrop for ontology building. Here are West and colleagues on these points:

We begin by describing a central challenge facing the study of addiction: the need to achieve clarity of constructs and develop consensus while at the same time recognising that divergent views have utility.

We can simply say more about what this means. On my view, this means Jim is disposed towards certain types of behaviors (including so-called ‘basic actions’ like intending, deciding, and so on) related to seeking and consuming heroin.

³³ An objector might wonder whether this is consistent with people being addicted to something that does not exist or that they have never come into contact with, such as an as-yet-undiscovered but highly addictive substance on Mars. If the reader is not worried about such science fiction scenarios, they should feel free to continue on. My view is consistent with this possibility. Technically, were the individual properly disposed, they would be addicted to *certain types of behaviors* related to that substance. ‘Jim is an addict’ is very much a fact (when it is a fact) about Jim and his properties, and so it is no problem accounting for Jim’s being addicted to some substance on Mars. We just need to remember that ‘Jim is addicted to Martian substance X’ is just an abbreviated way of speaking.

³⁴ This phrase is taken from West et al. (2019a) and is used in roughly the same way.

³⁵ In Chapter 4, I will explain how the *dispositional account* of addiction in particular can satisfy these desiderata. In Chapter 5, I will provide two more detailed sets of criteria (one substantive and one methodological, explained in that chapter) that will (i) serve as the basis for my comparative analysis of extant views (vis-à-vis the dispositional account), and (ii) still account for the three desiderata introduced in this section.

³⁶ My discussion here draws from the ideas found in Robert West and colleagues’ chapter in *The Routledge Handbook of the Philosophy and Science of Addiction* (West, Christmas, et al., 2019a).

We then move on to describe some key characteristics of the Semantic Web, and the ways in which these provide a pragmatic way of responding to this challenge.³⁷

I agree with West and his colleagues on three basic points that their work is getting at:

- (i) the addiction research challenge is a real problem for the study of addiction;
- (ii) that it centers around interdisciplinary dialogue and thus conceptual clarity; and,
- (iii) that ontology is the most viable way to solve this challenge.

However, West and colleagues further contend that, “While there is a need for greater clarity of constructs in the study of addiction, no investigator or organisation has the authority, or expertise, to propose a single unifying conceptual framework.”³⁸ Here, I believe they are mistaken.

It is not the case that some special authority or expertise is needed to offer a unifying conceptual framework. Still, it is a fair point that each addiction researcher has much to learn from the many others, whether within or outside of their home discipline.³⁹ Addiction is an extremely interdisciplinary field. What is more, perhaps partly due to a contemporary infatuation with the hard sciences (and brain imaging, in particular), much of the research on addiction is empirical – usually either clinical or neurobiological.⁴⁰ Hence, as with any science-dominated research area, a philosophical approach to addiction like mine needs to be sufficiently empirically informed, and so it needs to depend on the sciences involved. Yet it is also often quite easy to see that the various scientific and clinical approaches to addiction are in need of more careful and rigorous theorizing, and so depend on philosophy.⁴¹ Nonetheless, despite the interdependence of the various research perspectives on

³⁷ West, Christmas, et al. (2019a, p. 160).

³⁸ West, Christmas, et al. (2019a, p. 163).

³⁹ Indeed, I find this to be especially clear from a philosopher’s perspective.

⁴⁰ As a side note, a large majority on addiction are funded by the National Institute on Drug Abuse (NIDA), an organization that explicitly adopts the brain disease model of addiction (N. Volkow, 2014). I think that this inherently poses a problem, since it seems to potentially add to the disunification by placing much of the data we have on addicts’ and their brains within a particular and often narrowly construed paradigm about the nature of addiction. I come back to this point later.

⁴¹ One might think that I should say that the sciences depend on the humanities more generally (including history, sociology, critical theory, etc.) as opposed to just philosophy. I disagree. I do intend to use ‘philosophy’ here somewhat broadly, referring to its central tools and methods like conceptual analysis, theory building, critical thinking, and logic and argumentation. Perhaps these are taught and implemented across the humanities. Even so, it is philosophers whose job it is to master them. These are not add-ons that philosophers come across and might put into their toolkit. These *are* the

each other, I maintain that it is possible to propose a framework that can unify the literature in a substantive and meaningful way. What is needed is an approach which consists in utilizing a consensus terminology in a framework that can be used by biologists, psychologists, neuroscientists, and any other addiction researchers. Indeed, this is part of my very project.

With this in mind, I now turn to the need for a general account of addiction. I argue that the literature suffers from numerous problems, and that these are all problems of what I call ‘unification’. I then briefly motivate the claim that ontology is the proper (methodological) solution to such problems of unification. Before doing so, I lay out three desiderata for a successful general account of addiction. Lastly, I motivate the claim that the dispositional account defended here can serve as a single unifying framework.

1.3.1 Disunification in the Addiction Literature

There is no shortage of theories of addiction, nor of views about what is involved in addiction: genetic predispositions; environmental predispositions; underlying psychological mechanisms driving choice; underlying mechanisms in the brain; socioeconomic influences; psychological profiles and characteristics; physiological adaptations; natural histories, stages, or cycles; characteristic harms; comorbidities; causes and patterns of recovery and relapse; changes in self-identity; and on, and on. Perhaps surprisingly, this in itself is not the problem.

It is not a problem to disagree about a research topic, nor to approach it from different research perspectives. The key to successful research, especially regarding a highly interdisciplinary subject requiring significant collaborative efforts, is successful communication and integration of data. Disagreements and varied approaches do not in themselves preclude successful communication. The problem arises when communication breaks down. For instance, it is a problem when disagreement

philosopher’s toolkit. One might say that they *are* philosophy. Hence, while history and sociology and the like are important for addiction research, they do not undergird the sciences in the same way that philosophy does.

only *appears* to occur but is ultimately accounted for by interlocutors talking past one another. It is also a problem when disagreement occurs, yet the interlocutors fail to see why or what it is about. Another problem is the siloing of research efforts that (often inadvertently) makes interdisciplinary communication exceedingly difficult at best and either lost in translation or a non-priority at worst.

Addiction research suffers from all of these problems. However, it is unsurprising that it does, and it is not alone in doing so. These problems will plague the study of any complex phenomenon that spans multiple disciplines, such as well-being, love, religion and spirituality, intelligence, criminality, health, free will, and the like. It is hard enough to be clear and precise about what we mean when we communicate. This task is only made more difficult as the complexity of the topic of discussion increases, and so it is no wonder that such areas of research would face problems of communication. If we add to this the conceptual confusions and stalled debates that pervade the literature, the plethora of problems adds up quickly.

Let us turn to some examples from the addiction literature that illustrate its problematic state of disunification.

1.3.1.1 Vague or Unclear Definitions of Terms

One problem is that some authors are unclear about what their terms mean. This is sometimes due to their definitions or descriptions being incomplete, vague, ambiguous, or otherwise unilluminating.⁴² Consider George Ainslie's attempts at characterizing addiction:

We naturally overvalue the imminent future, and learn the ability to maintain long-term plans only gradually and imperfectly. Everyone struggles with bad habits. If addiction is defined with a low threshold, half the people in America are addicted to something (Sussman et al. 2011). Those of us who have avoided the named addictive diagnoses are nevertheless apt to suffer from habitual overvaluation of the present moment, as in chronic procrastination, overuse of credit, or unrealistic future time commitment. So the problem for the science of addiction is not an addict's susceptibility to temptation, but why she fails to use her culture's shared knowledge to counteract it in specific areas over part of her life...⁴³

⁴² It is also possible that these are *effects* of the authors being unclear about their terms. Either way, the point is the same.

⁴³ Ainslie (2019, p. 37).

A number of questions remain. Does Ainslie think addiction is no different from ordinary lapses in self-control? Is overvaluing the imminent future even an instance of a lapse in self-control? Does he think addiction is a *habitual* or *excessive* amount of such ordinary behavior? If so, what is the threshold for habitual or excessive? Can we define the threshold any which way, as Ainslie starts to suggest? Or is addiction indeed distinct from ordinary behavior, involving the absence of an ordinary ability to use shared cultural knowledge to thwart would-be failures of self-control, as Ainslie also suggests?

What makes his characterization even more unclear is that elsewhere Ainslie describes addiction in still further ways. In these further descriptions, there is reason to think that the inability to square one's actions with their long-term goals is central to his understanding of addiction:

Addictions are not simply recurrent impulses, but complex compromises with your long-term interests that develop when you try repeatedly to resist a temptation and fail.⁴⁴

An addict is not insensitive to differential motivation; her long-range interests just cannot get adequate leverage.⁴⁵

Still, though, we need more information. Ainslie himself points out that everyone suffers from this inability to act on their long-term interests at least sometimes. So, his definitions fail to tell us whether there is a difference between addicts and non-addicts on his view. Moreover, the above quotes suggest that Ainslie thinks addicts need to have had some particular history of failed attempts to control desires in order to have an addiction in the first place. It is unclear if this historical condition is intended.

It is hard to say whether Ainslie is providing an unclear definition or whether his account is unclear because he (perhaps unintentionally) provides multiple definitions.⁴⁶ It is also hard to say which is worse, but both are problematic enough for this to be beside the point.

⁴⁴ Ainslie (2017, p. 237).

⁴⁵ Ainslie (2017, p. 241).

⁴⁶ An objector may worry that Ainslie is providing neither a definition nor necessary and sufficient conditions but is instead merely characterizing addiction. First, it seems implausible that he is not trying to spell out the conditions that are central to his account of what addiction is – one quote even starts “Addictions are...” Second, even if this were true, it would be no better for Ainslie or the literature. He would simply have to be moved further down in this section where I discuss authors that provide no definition at all. Plenty other examples of vague and otherwise unclear definitions, conditions, and characterizations exist in the literature. Another objector might worry that it matters not what Ainslie intended or what his view suggests, but only what the most charitable way is to shape his view into a single, clear form. First, for the present

Gene Heyman, who follows Ainslie’s behavior-focused approach, provides characterizations of addiction that make Heyman’s understanding of the target phenomenon hard to pin down:

Individuals make choices according to quantifiable behavioral principles. Depending on specifiable conditions, these principles produce optimal outcomes, near optimal outcomes, or *seriously sub-optimal outcomes, which involve compulsive-like, excessive levels of consumption of a highly preferred substance or activity*.⁴⁷

Addiction is ‘disease-like’ in the sense that it persists even though on balance its costs outweigh the benefits... voluntary behavior that predicts the persistence of activities that from a global bookkeeping perspective (e.g., long-term) are irrational. That is, addiction is not compulsive drug use, but it also is not rational drug use... the defining features of addiction, which is to say its destructive and irrational aspects.⁴⁸

Addiction is ambivalent drug use, which eventually involves more costs than benefits.⁴⁹

Voluntary action and addiction differ in degree, not kind. For example, relapse and attempts to quit using drugs are signs of ambivalence, addiction by definition means excessive drug use...⁵⁰

Putting these together, on Heyman’s view addiction seems to be excessive, compulsive-like, voluntary, ambivalent, irrational consumption of a highly preferred substance that is on balance harmful. Is this any clearer? Perhaps, but only if we are clear about what all of those terms mean – what does ‘excessive’ or ‘compulsive-like’ come to? We also need to be clear about whether a history of actual consumption is necessary, as well as whether the addiction is identified with (or perhaps within) the principles of choice (such as hyperbolic discounting or melioration theory), their outcomes (such as choice patterns), or the outcomes of those outcomes (such as behavioral patterns or their consequences).

Nora Volkow – a neuroscientist and director of the National Institute on Drug Abuse (NIDA), and a leading defender of the *brain disease model of addiction* (BDMA) – also invokes different definitions in different contexts.⁵¹ In some cases, as expected, the account is neurobiological:

point of this chapter, this is incorrect. I am arguing that the literature is disunified, and Ainslie’s unclear and difficult-to-pin-down account is part of the evidence. Hence, his unclear and seemingly changing views are directly relevant. Second, in Chapters 4 and 5 where I show how the dispositional account unifies existing accounts and then evaluate extant accounts in the literature, respectively, I do just what is suggested. I narrow in (as charitably as I can) on a firmer view for Ainslie.

⁴⁷ Heyman (2019, p. 23), emphasis added.

⁴⁸ Heyman (2013, p. 1).

⁴⁹ Heyman (2013, p. 4).

⁵⁰ Heyman (2009, p. 124).

⁵¹ The BDMA is largely a neuroscientific view of addiction. The central thesis of the BDMA is that *addiction is a brain disease*, and more specifically, that addiction is a *chronic and relapsing* brain disease which is *developed as the result of continued use* and is *characterized by continued use despite harmful consequences* (N. Volkow, 2014). I discuss the BDMA more in later chapters. One

[Addiction is] a conditioned response [to the exposure] to the drug and/or drug-related stimuli that activates [the striato-thalamo-orbitofrontal] circuit and results in the intense drive to get the drug (consciously perceived as craving) and compulsive self-administration of the drug (consciously perceived as loss of control).⁵²

Addiction can be viewed as a pathology in how importance is attached to stimuli that predict drug availability and how the brain regulates (chooses) behavioral output in response to those stimuli.⁵³

Other times, however, Volkow's account is more psychological or straightforwardly appeals to (non-neurobiological) symptoms described within the DSM criteria:

Among the most insidious characteristics of drug addiction is the recurring desire to take drugs even after many years of abstinence... [and] the compromised ability of addicts to suppress drug seeking in response to that desire even when confronted with seriously adverse consequences... The enduring vulnerability to relapse is a primary feature of the addiction disorder.⁵⁴

Addiction: A term used to indicate the most severe, chronic stage of substance-use disorder, in which there is a substantial loss of self-control, as indicated by compulsive drug taking despite the desire to stop taking the drug. In the DSM-5, the term *addiction* is synonymous with the classification of severe substance-use disorder.⁵⁵

This problem of lacking clarity is exacerbated when authors fail to provide any definition at all. This sometimes occurs for 'addiction' itself, but also for addiction-relevant terms like 'control', 'compulsion', 'ability', 'disease', and so the like. For instance, it is exceedingly difficult to find a definition of 'disease' in any of Volkow's work on addiction (I have tried unsuccessfully to do so), and yet she is likely the most prominent defender of the BDMA (where 'D', after all, stands for 'disease'). Volkow's home organization NIDA also has no definition of 'disease' in its glossary on their homepage – the closest they come is in defining the DSM's label 'substance use disorder' as follows:

A medical illness caused by disordered use of a substance or substances.⁵⁶

might object that authors often invoke different definitions in different contexts outside of a discipline like ontology that is strictly in the business of crafting consistent sets of definitions. If so, then so much the worse for these authors, especially if they have any interests in effectively communicating with others about their view. Moreover, adding the need to search, merge, or analyze data (as in addiction research), which would need to be tagged with a vocabulary, only makes this worse.

⁵² Volkow & Fowler (2000, p. 323).

⁵³ Kalivas & Volkow (2005, p. 1410).

⁵⁴ Kalivas & Volkow (2005, p. 1403). Note that, according to this definition, addiction is a kind of term. This should sound odd since, like me, you probably took Volkow and others to be interested in addiction *qua* worldly phenomenon rather than 'addiction', the term we use to refer to that phenomenon. We will come back to this issue (called the 'use-mention error') in the next chapter's discussion of conceptualist vs. realist ontologies.

⁵⁵ Volkow et al. (2016, p. 364), bold in original.

⁵⁶ NIDA (2020). An objector might worry that this is unfair. The thought is that criticizing addiction researchers like Volkow or addiction research organizations like NIDA (both of which define 'addiction' in terms of disease) for not

Unfortunately, this is not very illuminating.

We will come back to the problematic conditions invoked by these authors just discussed. For now, consider only how difficult it is for other addiction researchers to understand a given account of what addiction is when it is entirely unclear what the author of the account is referring to when they use the term ‘addiction’, or whether they are always referring to the same thing. Is addiction in the brain (neural states or pathways), in the mind (desires, wants), in the external world (symptoms, behaviors), or some combination of these (whatever that would mean)? It is unclear whether a single author’s definition(s) could provide clear and consistent answers to these questions, let alone allow them to get on the same page with other researchers in answering them. As noted, the complexity of the phenomenon makes the latter task all the more important, and the failure to accomplish it all the more unfortunate.

1.3.1.2 Potayto, Potahto: Term Usage Differs between Authors

Aside from *lack of clarity* in definitions, another problem is that definitions of ‘addiction’, ‘disease’, ‘control’, ‘compulsion’, and the like *differ from author to author*. It is easy to see that innumerable accounts of addiction exist in the literature; but differing accounts of addiction-relevant concepts abound as well. Take ‘disease’, for instance. Jerry Wakefield, originator of the *harmful dysfunction account* of disease, defines ‘disease’ as follows:

...a medical disorder requires both dysfunction – that is, failure of some mechanism to perform a function that it was biologically designed to perform (where 'biological design' is understood in terms of the natural selection of evolutionary functions) – and harm, where the dysfunction causes harm to the individual as judged by social values.⁵⁷

having a definition of ‘disease’ would be on a par with criticizing me for not providing a definition of ‘philosophy’ at the start of the dissertation. However, these are disanalogous criticisms. Requiring that I define ‘philosophy’ would be analogous to my requiring Volkow and NIDA to define ‘medicine’ or ‘medical research’. I make no such claim. My claim is that ‘disease’ should be defined because it represents a central entity type that Volkow and NIDA talk about, study instances of, and define ‘addiction’ in terms of. The appropriate analogy would be to require me to define the entities central to my research (not the type of research itself), such as addiction, control, and desire. I do this.

⁵⁷ Wakefield (2017a, pp. 40–41). See Wakefield (1992, 2014) for a fuller treatment of his harmful dysfunction account.

Wakefield uses ‘disease’ and ‘disorder’ interchangeably, highlighting instead the distinction between *dysfunction*, on the one hand, and *harmful dysfunction* (disease or disorder) on the other. Evolutionary history obviously plays an important role in Wakefield’s account since it determines what the functions (and thus dysfunctions) are.⁵⁸ In the philosophy of medicine literature, Wakefield’s account is typically pitted against its most well-known rival, Christopher Boorse’s *biostatistical model* of disease.⁵⁹ While Boorse is not commonly known as an addiction researcher, his account of disease has been invoked by other addiction researchers in their discussion of the issue.⁶⁰ More importantly for the present point, Boorse’s account of disease differs from Wakefield’s, and so the two circumscribe different sets of phenomena when identifying diseases.⁶¹ For instance, if everyone in the world became an addict, this could not be a disease on Boorse’s biostatistical model but still could be on Wakefield’s harmful dysfunction account. Hence, some disagreements about whether addiction is a disease might come down to disagreements about what a disease is.

Now, this would not be so bad if it were more explicitly noticed. Unfortunately, it is uncommon for addiction researchers, especially those entrenched in clinical and experimental contexts, to invoke competing conceptions of ‘disease’ from the philosophy of medicine literature.⁶² What happens more often is that competing conceptions of ‘disease’ are simply taken on board without being fully fleshed out, and addiction is simply discussed within that framework.

Gene Heyman, for instance, paints a different picture of disease altogether. He explains (and seemingly endorses) what he sees as the BDMA view of disease as *involuntary behavior that has a biological, neurobiological, or genetic basis, and is medically treatable*.

⁵⁸ I define and discuss ‘function’ more in Chapter 3 where I introduce dispositions. For now, the point is again simply about providing evidence for the disunification of the addiction literature.

⁵⁹ See Boorse (1977, 2014).

⁶⁰ An example is Foddy (2011).

⁶¹ The direct exchanges between Wakefield and Boorse highlight this point (Boorse, 2014; Wakefield, 2014).

⁶² Wakefield, in particular, is a welcome exception to this rule, making conceptions of ‘disease’ very explicit in his exchanges with other addiction researchers (Wakefield, 2017a, 2017b).

...scientific explanations for why addiction should be considered a disease depend on assumptions regarding the relevant categories, which in this case are involuntary and voluntary behaviors. For example, one of the mainstays of the claim that addiction is a disease is evidence that it has a genetic basis. The idea is that if genes influence an activity, then it can't be voluntary... if a key feature of a disease state is that the symptoms are involuntary, then we need to know how to distinguish between voluntary and involuntary behavior.⁶³

...there are also important empirical arguments to consider [that] typically focus on three lines of evidence and reasoning: (1) addiction has a biological basis; (2) addictive drugs have the capacity to transform a voluntary user into an involuntary one; and (3) the disease interpretation leads to better treatment for addicts.⁶⁴

Heyman has yet another distinct understanding of *disorder*:

...to say that addiction is a disorder is to say that it is not an optimal pattern of behavior.⁶⁵

Given his fuller account, 'non-optimal' here seems to mean either that the behavior produces (or, more charitably, *tends* to produce) harmful consequences, that it is irrational (by conflicting with one's long-term goals), or both.

Differing definitions of 'addiction' and 'disease' are plentiful enough to facilitate substantial disunification, confusion, and misunderstanding. The problem widens, though, as one considers further addiction-relevant definitions of 'control', 'compulsion', 'harm', 'significant', and so forth.⁶⁶ Control is sometimes understood as *reason-responsiveness* – roughly, the capacity to recognize and react to reasons for and against one's choice options.⁶⁷ Other times it is cashed out in terms of willpower.⁶⁸ Compulsion is often the subject of dispute since some hold that it entails *irresistibility* while others deny this.⁶⁹ By now the point is hopefully clear. Even when authors make it sufficiently easy to identify their definitions, there is still no real shared vocabulary or understanding of the phenomena in question.

⁶³ Heyman (2009, pp. 89–90).

⁶⁴ Heyman (2009, p. 91).

⁶⁵ Heyman (2009, p. 124).

⁶⁶ Note that this is true even if for many (or even most) actual cases the authors' views are co-extensive. If they understand 'disease' or 'addiction' differently, they are inevitably going to talk past one another, spilling more and more wasted ink. Moreover, the problems that such incompatible definitions create for doing ontology would also still exist. More on ontology in the next chapter.

⁶⁷ Some addiction researchers explicitly speak in terms of reason-responsiveness (Levy, 2015; Sinnott-Armstrong & Pickard, 2013; Uusitalo, 2011). See Fischer and Ravizza (2000) for the original account of control as reason-responsiveness.

⁶⁸ See Wallace (1999) for an example.

⁶⁹ Many examples exist for both views, compulsion as irresistibility (Andreou, 2007; Benn, 2007; Frankfurt, 1971; Herdova, 2015; Pickard, 2015; Watson, 1977) and compulsion without irresistibility (Charland, 2012; Heather, 2017c; Henden, 2013a;

1.3.1.3 Some Definitions Are of Low Quality

Yet another problem is the *quality of the definitions* provided (or implicitly assumed) by some authors. Consider Heyman's definition of 'disorder' given above. Whether he intends it to mean *harm-producing* or *irrationality*, the definition is problematic. Mixed martial arts (MMA) fighters engage in behavioral patterns with systematically harmful consequences, yet this does not make their behavior disordered. Perhaps Heyman would say that MMA fighters' behaviors are never *on balance* harmful. This only introduces new problems since, aside from that being a questionable empirical claim, it would mean that many apparent addictions are either not disorders (contrary to his book title and central claims) since they are not on balance harmful, or they are not addictions (if harm/disorder is necessary for addiction, as Heyman implies). Either way he would have *even more* explaining to do.

Furthermore, the definition is no better off if 'disorder' means that the behavior is irrational. People can behave irrationally without the behavior being disordered. It may be irrational for me to quit my job out of anger at my boss, but this does not mean I am disordered. In fact, *akratic* actions – roughly, acting against one's better judgment – are typically understood to be free actions, and so being able to act contrary to one's reasons may be part of what it is to act freely.⁷⁰ Depending on one's understanding of rationality, conspiracy theorists and even some extremely religious individuals might systematically act irrationally. Yet this does not necessarily make them disordered. We also have systematic psychological biases, some of which may count as irrational. For instance, people habitually make errors regarding probability and statistics or the behavior of physical objects, even though they should have enough evidence to avoid these errors.⁷¹ However, they are not disordered, and in fact

Henden et al., 2013; Hyman, 2007). See Heather (2017c) for an excellent discussion of the varying uses of 'compulsion' in addiction research and their respective utility (or disutility).

⁷⁰ See Mele (1992, 2008) and Franklin (2015) for examples.

⁷¹ See Kahneman & Tversky (1979) and Kahneman (2013).

some of these biases seem to be byproducts of quite helpful, well-designed mechanisms of thought that, on the whole, work rather well (for instance, for reducing cognitive load).

Consider further Wakefield's definition of 'harm'. Wakefield's account of addiction nicely distinguishes *addiction* (a certain dysfunction of an agent's desire/deliberation/choice process) from *addictive disorder* (an addiction that is harmful) since, on his view, diseases/disorders are *harmful dysfunctions*. But the sense of harm he invokes is based on social values.⁷² In other words, whether a dysfunction is harmful (and thus a disease) depends on the values of the society the individual belongs to. If this were right, diseases could be cured by flying to a country with wholly different values and immersing oneself in their culture. Alternatively, the society as a whole could change their values, curing many diseases in doing so. What about non-human (or non-social) organisms? Is a chimpanzee's cancer a disease only according to the values of the chimp's social group? What would that mean? It seems even less desirable to say that the chimp's dysfunction is a disease only if it is harmful according to some *human* society's values. Moreover, if either epicureanism or theism is true, then death is not (necessarily) harmful or a bad state, and thus some fatal conditions would not be diseases.⁷³ To be clear, the problem is not that Wakefield's account of disease is normative. It is that the notion of harm invoked is unsatisfactory.⁷⁴

Finally, consider some examples of how 'control' gets defined. As noted, some authors, to their credit, invoke more sophisticated conceptions of control like Fischer and Ravizza's reason-responsiveness. However, others claim that addicts lack control in the sense that they are *compelled*, where this seems to mean that they *literally cannot do otherwise*. But this would mean that if they were able to do otherwise, then they would not be compelled and would therefore be in control. Indeed,

⁷² Wakefield (1992).

⁷³ Hershenov (2020).

⁷⁴ See Feit (2017) for an argument that no view of harm will save Wakefield's account. See Limbaugh (2019) for a response to Feit and a defense of a modified version of Wakefield's normative account that he claims survives Feit's objections. However, see Feit (2020) for a response to Limbaugh.

this is precisely the typical (and unsurprising) response from critics of these authors. The problem here is that these authors seem to confuse *the exercise of a capacity* (or failure to exercise it) for *the capacity itself* (or for its absence). Lack of control, whether instances of it or the lack of a general capacity, does not require literal irresistibility (nor does compulsion). Nor do instances of literal irresistibility entail lack of a general capacity for control. And instances of control or choice do not entail the presence of a general capacity for control. Others seem to define ‘control’ or ‘compulsion’ in terms of harmful consequences – that is, to continue some behavior despite harmful consequences is to be compelled.⁷⁵ However, harmful consequences are either irrelevant to control altogether or they are merely evidence for impaired control rather than what impaired control is.

1.3.1.4 Interpreting the Evidence for Disunification

Plenty more examples of these kinds of conceptual confusions and definitional blunders exist, but hopefully the point has been made. Many of the central concepts surrounding addiction (including ‘addiction’ itself) are often undefined, vaguely or ambiguously defined and characterized, defined in multiple ways (by the same or different authors), or poorly defined. The issues canvassed here all reflect problems with what I call ‘unification’. Methodological unification is *the extent to which researchers studying a given topic have adopted both a shared understanding of the target phenomenon and a shared approach to achieving that understanding*. Put another way, it is the extent to which researchers in a domain are talking about the same thing, and the extent to which they can and do work together. Addiction research currently lacks sufficient unification, as exhibited by the numerous problems above.

To be sure, methodological unification does not require absence of disagreement. It only requires that, when disagreement occurs, it is genuine and recognizable. This is accomplished through, among other things, having a shared vocabulary through which to express the disagreement. If one

⁷⁵ Proponents of the brain disease models are notorious for this. See almost any of the numerous brain disease accounts referenced in the footnotes above.

person claims that addiction is a disease, for instance, while another claims that it is not a disease, then ‘addiction’ and ‘disease’ better mean the same thing to each person. Otherwise, not much progress will be made. Additionally, lack of unification does not require absence of *attempts* to dialogue or work together. This is because insufficient methodological unification does not only hinder progress, but also the ability to see *that* (or *why*) progress is hindered. Hence, efforts of dialogue and collaboration may continue in an insufficiently unified research domain.

1.3.1.5 Two Consequences: Shoddy Work Rolls Downhill

There are two practical issues that these problems create, further adding to the disunification in the literature. First, the theoretical and conceptual issues just canvassed can and do trickle down to clinical and experimental contexts. For instance, consider the debate about whether addicts should be allowed to take part in clinical trials involving consumption of their drug of choice since their autonomy (and hence, their consent) may be compromised.⁷⁶ This inevitably involves questions concerning whether, when, and how addicts might suffer autonomy impairment, compulsive desires, and the like. More importantly, it also requires unpacking these terms and getting everyone on the same page. If this has not been done, as the above discussion demonstrates, it is unclear how the practical questions can be effectively discussed, let alone sufficiently answered.

Consider also how we determine who counts as an addict and a non-addict (or member of the control group) in such trials, as well as any other empirical or experimental data on addicts. For such data, the DSM-V criteria (or that of whichever version was current) is used for inclusion into the *addict* group virtually across the board. But the DSM-V contains *diagnostic criteria*, and so is not necessarily intended to provide a definition of ‘addiction’.⁷⁷ At the same time, many authors do not treat it this

⁷⁶ See Henden (2013b) for this view. See Uusitalo & Broers (2015) for a contrary view. See Henden (2016b) for a reply.

⁷⁷ See Summers (2015). However, note that the introductory section in the DSM-V, explaining how the manual is to be used, is somewhat ambiguous on this point (American Psychiatric Association, 2013, p. 10, cf. p. 19).

way, which helps to explain the plethora of differing accounts. But if the DSM-V is the gatekeeper, deciding who and what counts as an *addict* or *addiction*, then the data we have on addicts and addiction may not be properly representative of the actual phenomena it is meant to represent or describe. This is because the DSM-V is at best only identifying *evidence* of addiction, and at worst only provides a way to identify *problem users* rather than addicts proper. Problem users are anyone who experiences personal or social problems resulting from some type of behavior (drinking, drug use, gambling, shopping) for which they feel they need professional help. Addicts are not necessarily all problem users and problem users are not necessarily all addicts. Hence, if the DSM-V helps identify problem users, then subjects selected by its criteria may include some non-addicts and may fail to include some addicts.⁷⁸

If data obtained or the knowledge based on it is used (explicitly or implicitly) to develop further empirical investigations and obtain further data, then the problem is only exacerbated. It can become a worrisome self-fulfilling prophecy. The upshot is that these problems arise because what gets done at the practical – that is, clinical or experimental – level requires applications of the theoretical and conceptual work that was (supposed to have been) done beforehand.

Second, setting aside these worries about clinical trials and study inclusion, consider the interoperability of the various kinds of data we have on addicts and addiction. Data is collected by historians, sociologists, social workers, nurses, epidemiologists, biologists, neuroscientists, psychologists, psychiatrists, and economists, among others. However, since unification is lacking both within and across these different disciplines (and their sub-disciplines), sharing and understanding each other's data must also lack sufficient unification. In other words, even if researchers can manage to avoid the problems just described and get their own good data on addiction, difficulties still remain concerning the interoperability of that data with research from other disciplines. This is due to the *siloining* of information. This is when information (or data) within one part of an institution or system

⁷⁸ A similar point has been made elsewhere (Sanislow et al., 2020, p. 64).

(like academia, or addiction research) fails to be interoperable with other parts it should be able to communicate with.⁷⁹

For instance, suppose neuroscientists at a university have data on consciousness or some aspect of it. Researchers in other departments like biology, psychology, or philosophy have an interest in this data. However, they may lack the requisite framework for interpreting the neuroscientific data. Alternatively, they may have an incompatible way of representing the same (or related) phenomena – their data may be modeled differently, using different terms, definitions, and so forth. The neuroscientific data would be siloed since it is not interoperable with what should be a related system of data collection or research. This can happen for several reasons, but important for our purposes is that the disunification laid out above is perfectly sufficient for facilitating information siloing. It is likely that this is often inadvertent, resulting merely from researchers in a given discipline or sub-discipline labeling and defining their data in a way that is easy, familiar, or useful to them and their particular aims. It is also possible that it results from a lack of concern for making one’s data interoperable with other disciplines. Nonetheless, information siloing is problematic whether it is inadvertent or not. Sharing and understanding data across disciplines and research programs is critical if the end goal is to have the best, most comprehensive understanding of the phenomenon in question.

1.3.1.6 A Quick Recap

A brief survey of different accounts of addiction in the literature demonstrates the absence of a shared understanding of the phenomenon in question. Some descriptions are vague or ambiguous, making it hard to pin down what exactly addiction is supposed to be. Others provide either no definition at all or no definition of terms central to their definition, such as ‘disease’, ‘control’, or ‘excessive’. Some accounts provide more than one definition. Some definitions are flawed because

⁷⁹Adapted from https://en.wikipedia.org/wiki/Information_silo.

they confuse things like evidence, causes, and influences of addiction with addiction itself. Moreover, when there is significant disunification at the theoretical and conceptual level, disunification at the level of application should be incredibly unsurprising. The central issue here is that there is no shared understanding of the phenomenon in question, no shared vocabulary for discussing it, and, perhaps most importantly, no common methodological ground from which the problem can be addressed. In short, there is insufficient methodological unification.

1.3.2 Ontology as the Solution to the Disunification Problem

How do we go about solving a disunification problem like the one laid out above? The present section answers this question, while leaving the fuller discussion of the methods and principles, and applying them to addiction, to the following chapter. The short answer is that ontology helps to solve this problem. Here, I motivate this solution by explaining the reasoning behind it.

A handful of researchers working on addiction and familiar with the methods of ontology have called for the latter's introduction into the domain of addiction research. This is because addiction research is not the only field to have experienced the level of disunification described in the preceding section. Moreover, and fortunately for addiction research, one method of solving this disunification problem has proven to be hugely successful in other domains. That method is the incorporation of realist ontologies, built using the principles of best practice underlying Basic Formal Ontology (BFO) and adopted by the Open Biological and Biomedical Ontology (OBO) Foundry.⁸⁰

1.3.2.1 Realist Ontologies and the Semantic Web

A realist ontology is a *controlled, structured vocabulary the terms of which refer to portions of reality*. The terms in an ontology are defined using ordinary language so that humans can understand them, but also given formal definitions so that their logical structure is laid bare in order for computers to reason

⁸⁰ See (Arp et al., 2015; Ceusters & Smith, 2010; The OBO Foundry, 2020b).

over them. Ontology makes use of the Semantic Web.⁸¹ This can be thought of as an advanced version of the World Wide Web which, in addition to textual content (strings of text), contains *meanings* (semantic content) for each piece of information defined in terms of *triples*. Triples are subject-predicate-object statements that identify a term's location in the ontology. Recall that an ontology starts with a hierarchical taxonomy of entity types (universals or classes) and their type-subtype relations as their backbone. So, the triples tell you what the term means (what it refers to) and how it is related to other terms (representing other kinds of entities) in the hierarchical structure.

For instance, **PERSON** *is_a* **ORGANISM** tells me that every instance of **PERSON** will possess the essential features that all instances of **ORGANISM** share (metabolism, growth, and so on). But there are many subtypes of **ORGANISM**, meaning that this more general type can be divided into still further subtypes of organisms based on shared, essential features that distinguish them, such as **PLANT**, **MAMMAL**, or **SLUG**. For instance, suppose 'person' is defined as *an organism that can reason and has a soul*. If so, then **PERSON** *is_a* **ORGANISM** would also tell us that instances of **PERSON** (such as you and me) are not only organisms, but also reasoners with souls. Of course, there are many other features or properties that persons have, such as in virtue of also being instances of **MAMMAL**, **MALE**, **STUDENT**, **SON**, and so on. Moreover, they stand in numerous relations beyond the *is_a* relation, such as having parts (*has_part*), participating in processes (*participates_in*), and so on. These additional properties and relations that are relevant to what persons are, and how persons are related to whatever kinds of entities they are related to, can be added into the ontology. We need only define the other kinds of entities, properties, and relations in a similarly clear and logical way.

The Semantic Web, then, helps to organize all of this information about the domain of interest being represented. Once defined, each entity is assigned a *uniform resource identifier* (URI) that tells the computer where the entity is in the ontology by way of formal definitions and logical relations. These

⁸¹ The following discussion of the Semantic Web draws from West, Christmas, et al. (2019b, pp. 164–165).

can then be used by the computer to effectively search the ontology. The use of triples that lay bare the logical structure of our definitions forces clarity and uniformity of constructs.⁸² Moreover, the URIs make it so that meaning comes first, not labels. It matters not what label is used for a particular entity once defined into its place in the ontology. Instead of ‘person’, we could have used ‘Marphlong’ to refer to *organisms that can reason and have a soul*. Whether we want to call Bob a ‘person’ or a ‘Marphlong’ is irrelevant. What matters is that we are clear about the definitions and relations associated with a particular URI, since these give us the fuller meaning of the kind of entity in question that is being represented. What we name it is neither here nor there.

1.3.2.2 Controlled Vocabularies Facilitate Interdisciplinary Research

When vocabularies are controlled in this way – clear, shared, computational, and with an eye towards realism – it is easy to see how utilizing them makes theory and data much more interoperable. Indeed, ontology has been widely successful in unifying research efforts and perspectives surrounding various fields of study, with the Gene Ontology (GO) being the best and most well-known example.⁸³

Here are Robert West and colleagues on this point:

To date, these tools have proved extremely valuable in other fields of social and behavioral science (Larsen *et al.* 2017) and clinical science (Lewis 2017), and there is reason to believe they could lead to more rapid advances in our understanding of addiction as well.⁸⁴

In calling for reform to the addiction literature, West, Janna Hastings, and others argue that introducing ontology building would likely be just as successful there as it has been in other areas. I think they are correct. The argument implicit in these authors’ works is as follows:

- (P1) Implementing the methods and principles of realist ontology building into domains like biology and medicine helped solve their problems of disunification;
- (P2) With respect to solving their problems of disunification, there is no relevant difference between these domains and the domain of addiction research in particular;

⁸² See West, Christmas, et al. (2019b, p. 164).

⁸³ See (Ashburner et al., 2000; The Gene Ontology Consortium, 2019). I will explain this success more in the next chapter.

⁸⁴ West, Christmas, et al. (2019b, p. 171).

- (P3) If (P1) and (P2), then implementing the methods and principles of realist ontology building into addiction research will help solve its problems of disunification;
- (C1) Hence, implementing the methods and principles of realist ontology building into addiction research will help solve its problems of disunification. [(P1)-(P3)]

Premise (P1) is well-supported by the success of ontologies like GO and the Ontology of General Medical Science (OGMS). Premise (P3) is trivially true. Thus, premise (P2) is the most contentious. However, this overstates things since, at least at face value, it is hard to see what relevant difference there could be between addiction research and other domains of research implementing ontology. That is, (P2) hardly seems contentious. Biology and medicine are both highly interdisciplinary (arguably more so with the help of ontology). Much addiction research even falls within one or the other of these domains. What is at issue with ontology and unification is a shared way of talking about the entities in question, and a shared methodology for organizing and using that vocabulary. Given that, it is hard to see how *any* domain of research would be relevantly different from those that have seen successful unification through the implementation of ontology. Therefore, (C1) looks to be very well-supported.

1.3.2.3 What a General Account Comes To

In light of the foregoing discussion, it is best to further clarify what I mean by saying a ‘general account’ of addiction is needed. Ontology allows for mutual or cross-disciplinary understanding of some phenomena being studied. In other words, it solves the problems of disunification because it provides the opposite. To somewhat oversimplify things, it does this by allowing researchers from any discipline or perspective to talk about the same thing, or to be completely aware when they are not. Clear, shared, explicit definitions for each kind of entity and relation make this possible. So, as West and colleagues point out, through the use of ontology we are able to solve the addiction research challenge by providing “a way to achieve clarity of constructs and develop consensus, while at the

same time recognising that divergent views have utility.”⁸⁵ This is what they mean, and part of what I mean, too, when we argue that addiction research needs a general model of addiction.

Consider again the elephant metaphor. The analog of implementing ontology would be something akin to giving each man his sight back, and then asking him to try providing an account of ‘elephant’ once again. If each man understood that he and the others studied one aspect of the same thing, and each knew which aspect was studied by whom because they *actually understood* each other’s descriptions and what they referred to, then theorizing about the elephant would go much more smoothly. We would have multiple perspectives effectively working together to converge on a general theory of elephants, rather than a plethora of views that lack the means to become integrated. Ontology introduces a way of developing a general model of addiction as described by West and colleagues. That is what is needed since *that is what it is* to contribute to a more unified literature.⁸⁶

It should be clear that this is indeed the proper first step to take in solving the problems of disunification the addiction literature is suffering from. In the next section, I identify three desiderata for a satisfactory general account of addiction.

1.3.3 Desiderata for a General Account of Addiction

To say that the literature needs a new and better account of addiction, and a new and better method underlying it, is to implicitly say at least two things.⁸⁷ First, it implies that an account of

⁸⁵ West, Christmas, et al. (2019b, p. 163).

⁸⁶ While the endeavor is new and yet underdeveloped, the implementation of ontology into the addiction literature is gaining some observable traction. West and colleagues’ paper calling for the use of ontologies in addiction research appeared in the most recent *Routledge Handbook on the Philosophy and Science of Addiction* (Pickard & Ahmed, 2019). Susan Michie and colleagues at University College London’s Centre for Behavior Change are working on the collaborative Human Behavior-Change Project, including the development of the Behaviour Change Intervention Ontology (BCIO): <https://www.humanbehaviourchange.org>. This is relevant because addiction involves behavioral changes and addiction therapies are behavior-change therapies of one form or another. Also recall that the journal *Addiction* recently announced a new series (called “Addiction Theories and Constructs”) meant to provide a shared forum for addiction researchers that will help to realize the aim of developing addiction ontologies. My project can also be seen as a contribution to that goal.

⁸⁷ To be clear, by calling my account ‘new’ I mean something like: (i) no one has yet explicitly argued for a dispositionalist account, at least not with the metaphysical detail that I offer; and (ii) despite the fact that many accounts are implicitly dispositional, no one seems to have recognized this. Moreover, calling ontology ‘new’ is restricted to its implementation into the addiction literature. It is not yet a method that has been accepted (or even acknowledged) across the literature.

addiction should be a certain way (or do a certain thing). Second, it is to say that existing accounts are not that way (or are not doing that thing, or not doing it well). The latter claim was motivated above and is addressed in detail in Chapter 5 when I compare my account to extant views. In this section, I will be concerned with the former claim. I first explain three desiderata concerning three broad issues: empirical data, theoretical breadth, and level of analysis. Then, I close the section with a brief discussion of the desiderata proposed by Robert West and colleagues.

1.3.3.1 Empirical Sensitivity

First, and perhaps most obviously, any account of addiction should be *empirically sensitive* to what we know about this phenomenon. There are at least three forms of empirical sensitivity that are relevant: consistency, explanation, and neutrality. First, one's account should be *consistent* with well-established and relatively uncontroversial data surrounding addiction. Consider some of the most widely known and agreed upon features of addiction. For instance, addiction seems to have various ways in which it can be manifested, such as the variation in things – that is, substances or behaviors – people can be addicted to (drugs, sex, food, water, gambling, shopping, texting, sexting, eating drywall, and so on).⁸⁸ A good account of addiction would allow sufficient variation in how addiction gets manifested since such variation is *prima facie* plausible, given the evidence. Similarly, a good account of addiction would allow for other well-known and commonly accepted features of addiction, like the fact that addicts exhibit exaggerated hyperbolic delay discounting,⁸⁹ often eventually age out of addiction without treatment,⁹⁰ appear to sometimes make choices,⁹¹ and typically undergo certain

⁸⁸ Differing manifestations of addiction (in this sense) have been well documented (American Psychiatric Association, 2013; Foddy, 2017; Heather, 2017b; Hutcheon & Bevilacqua, 2010; Karila et al., 2014; Meule & Gearhardt, 2014; Robbins & Clark, 2015). Eating drywall is only one of some very strange addictions that have been reported: <https://www.discoveryuk.com/series/my-strange-addiction/>.

⁸⁹ Ainslie and his colleagues have been central in showing this (Ainslie, 2001, 2013, 2017, 2019; Ainslie & Monterosso, 2003; Bickel et al., 1999; Bickel & Marsch, 2001). For more detail on delay discounting, see Appendix A.

⁹⁰ Gene Heyman has often made this point in critiquing brain disease views (Heyman, 2009, 2013).

⁹¹ This point, too, is often made during criticisms of the brain disease view, given its overly restrictive conception of control (Heyman, 2009, 2013; M. Lewis, 2015; Schaler, 2000; Sinnott-Armstrong & Pickard, 2013; Uusitalo et al., 2013).

characteristic brain changes.⁹² At minimum, these should all be consistent with one's account. Given their plausibility and empirical support, if they are not, the burden is on that account to explain why.

In addition to consistency, empirical sensitivity will also involve an *explanatory* component. More specifically, this involves the puzzle of addiction, which concerns addicts' continued engagement in their addictive behavior despite enduring significant costs and often expressing a desire to stop (or, at least, sincere ambivalence or regret). A good account should not just aim to be consistent with the fact *that* addicts behave the way they do, but it should also aim to explain *why* they do. Indeed, this is likely why some authors think that answering this question thereby provides an account of what addiction is. An account is better insofar as it can explain this puzzle – though *the account* is still needed, which is distinct from the explanation it provides of this puzzle. Where an account does not offer novel or improved-upon explanations, it should be consistent with the best existing explanations of such phenomena. This is where the account of addiction being *general* in the sense described above can come in handy. An account of addiction that utilizes ontology in its development will naturally lend itself to interoperability with different views.

A final type of sensitivity to empirical data comes in the form of *neutrality*. The thought here is that all else being equal, it will count in favor of an account of addiction that it is neutral with respect to a particular debate or point of contention where the evidence is less than conclusive, and controversy remains. For instance, there are various neurobiological models of addiction proposing different answers to the question of how addiction fundamentally works at the level of brain mechanisms.⁹³ However, it would seem to favor one view of addiction over another if, all else being

⁹² A popular point from brain disease views (N. Volkow, 2014; N. Volkow & Fowler, 2000; N. Volkow & Morales, 2015), though even their opponents accept this since brain changes do not entail brain disease (M. Lewis, 2015; Wakefield, 2017a, 2017b).

⁹³ Some are habit-based (Dalley et al., 2011; Everitt & Robbins, 2005, 2016), some are positive reinforcement models (Berridge & Robinson, 2016; Grodin et al., 2016; George Koob, 2013; Linden, 2012; Robinson & Berridge, 1993), some are based on opponent processes (Solomon, 1980), some are based in an allostatic process (G. F. Koob & Le Moal, 2006), and some are associative learning models (Arpaly & Schroeder, 2014; Di Chiara, 1999; Hyman, 2005; Redish, 2004).

equal, it was able to remain neutral with respect to which neurobiological model turned out to be correct (or, perhaps, which neurobiological mechanisms turned out to be most central). For instance, suppose an account of addiction committed its adherents to the *allostatic model* or the *opponent process model*, but a second account was consistent with both neurobiological models because it didn't require any *particular* brain states or processes be present in an addiction. If all else is equal – they explain the data equally well and so on – then it seems much better to remain neutral, since then the success of the view would not be hanging by a neurobiological thread, so-to-speak.

1.3.3.2 Theoretical Breadth

Second, an account of addiction should possess significant *theoretical breadth*. Here I mean that an account should be able to accommodate what is *possible* concerning addiction since the aim should be to understand *the nature* of addiction. For instance, if it is possible that a particular addict is never harmed, then an account of addiction should capture this. Otherwise, the account is either incomplete or not genuinely an account of addiction, since we would have a case of addiction that the account does not get right. I suspect that this is implicit in most addiction researchers' approaches, and in particular any scientific approach. They are all trying to get at some phenomena in the world – some *portion of reality* – and to get it at fully and accurately. Additional possibilities that ought to be accounted for might include: addicts without a history (whether a certain type or one at all); unmanifested (perhaps even asymptomatic) addictions; and multiply realizable addictions. If these are possible, a satisfactory account of addiction will capture them. Moreover, the aforementioned point about neutrality applies here. That is, it would be preferable for an account to remain neutral concerning any unresolved theoretical disputes, such as the debate over whether addiction is a disease.⁹⁴

⁹⁴ That is, it should be neutral when laying out the very nature of addiction. It may be possible to then turn straightaway to saying whether addiction turns out to be a disease (given one's account of disease). The point is just that these two accounts – of the natures of addiction and disease – come apart. Neither should be defined in terms of the other.

1.3.3.3 Proper Level of Analysis

Third, an account of addiction should adopt the proper *level of analysis*. More specifically, one should adopt a domain-neutral approach as far as possible, rather than one that is too narrowly focused on one particular disciplinary perspective. As we saw above, discipline- or domain-specific approaches and information siloing often go hand-in-hand and contribute to the numerous problems of disunification. Hence, the account should, insofar as is possible, be able to unify the various ways in which addiction is studied and talked about across disciplines. It should already be relatively clear that the ontological approach I adopt will point us in that direction (the next chapter explains *how* it does so). Moreover, as in the elephant metaphor I take there to be a common underlying phenomenon that addiction researchers are getting at – there is a *there* there. Implementing ontology as the methodological foundation of the account provides the domain-neutral approach we need to capture this. We will also come to see as the account is filled out in Chapters 3 and 4 that the dispositional nature of addiction is useful in this regard as well. Hence, both the approach *and* the substance of the account can help to remedy and avoid the disunification that hinders the literature.

1.3.3.4 How these Desiderata Capture those from West and Colleagues

Robert West and colleagues have discussed the desiderata of a general model of addiction as well, but the criteria above encompass their desiderata. In their paper calling for ontology in addiction research, West and colleagues argue that a satisfactory model of addiction ought to make sense of the diversity of views across the various disciplinary approaches to studying addiction. Further, it should respect the utility of each perspective for understanding addiction more fully and comprehensively:

A general theory of addiction has yet to be developed, but a key requirement for such a theory is that it should recognise and accommodate multiple viewpoints on addiction, and not be limited to a single viewpoint such as the ‘medical model’ (construing addiction in term of a mental disorder, disease or disease process)... On the other hand, it should go without saying that a general theory that achieved greater clarity by simply ignoring this diversity of viewpoints would not in fact be a general theory at

all. To move forward, we need a way to achieve clarity of constructs and develop consensus, while at the same time recognising that divergent views have utility.⁹⁵

As was just noted, the method of ontology provides generality to the account, and generality is part of what provides unification. The authors go on to explain that this would involve something like what I called ‘empirical sensitivity’ above. That is, allowing a diversity of views and perspectives means capturing the well-established empirical data, such as which motivational or neurobiological processes are involved. They provide a table displaying some of these processes, including reward seeking, attachment, incentive sensitization, impaired control, social influence, and so on:

Any general model or theory of addiction would need to capture processes that have been identified as important in its development and maintenance (Table 13.1).⁹⁶

Thus, the need for an account of addiction to be inclusive of various approaches, and to recognize their utility, is already captured by the three desiderata I provided above. However, perhaps one last desideratum should be identified that will make sense of West and colleagues’ (and my) intention to implement the principles and methods of ontology in approaching the study of addiction.

We share this intention because we share the view that the addiction literature is seriously disunified, and that ontology can help to provide the requisite unification. A central part of the reasoning behind this view is that ontology allows even extremely heterogenous disciplines and perspectives, most with extremely heterogenous datasets, to effectively communicate and share and manage that data. Hence, a good account of addiction should facilitate *interdisciplinary dialogue*, where this includes not only a way for various disciplines to meaningfully *communicate* (such as through a shared vocabulary), but also the *interoperability* (shareability, searchability, understandability, mergeability) of their data. Thus, putting all of this together, a good account of addiction will have the following features:

(i) **Empirical sensitivity:**

⁹⁵ West, Christmas, et al. (2019b, p. 163).

⁹⁶ West, Christmas, et al. (2019b, p. 161).

- a) *Consistency* with well-established data;
 - b) Provides or is consistent with the best *explanation* of data; and,
 - c) *Neutral* with respect to unresolved empirical disputes.
- (ii) **Theoretical breadth:**
- a) *Possibilia* captured; and,
 - b) *Neutral* with respect to unresolved theoretical disputes.
- (iii) **Domain-neutral level of analysis.**
- (iv) **Facilitates interdisciplinary dialogue:**
- a) Shared meaning and understandability: and,
 - b) Interoperable data (integrate results, test predictions).

I now turn to final point about the advantage of the substance of my account that challenges West and colleagues' claim that no single conceptual framework can be unifying.

1.3.4 A Dispositionalist Account as the Foundation for Unification

When West and colleagues call for the addiction literature to be cleaned up through the use of ontology, one might wonder why they then go on to claim that “no investigator or organisation has the authority, or expertise, to propose a single unifying conceptual framework.”⁹⁷ One might think, did you not just do that? Is ontology not the single unifying conceptual framework? As I understand them, the answer would be “no” because ontology provides a *methodological* framework. This is part of its appeal, since it allows for all types of content to be plugged into the machinery, helping to achieve the inclusion of diverse views, a feature they were looking for. Here is West and colleagues discussing some benefits of using the Semantic Web (part of the ontology machinery):

For example, different people can continue to use terms such as ‘addiction’ differently, since these terms will be assigned separate URIs and definitions despite sharing the same superficial label. In fact, there is a built-in way to represent different “perspectives” on the Semantic Web through use of different ‘namespaces’. Stated simply, if I want to define addiction in a new way I can simply create my new meaning of addiction in my own unique namespace, and explicitly refer to that version of addiction in my content. This makes it completely clear what the term means as I am using it. Importantly it also allows for systematic comparisons between my usage and someone else’s.⁹⁸

⁹⁷ West, Christmas, et al. (2019b, p. 163).

⁹⁸ West, Christmas, et al. (2019b, p. 161).

In other words, “*anyone can say anything about anything.*”⁹⁹ This is appealing because it allows diversity. However, it also explains their view that no single conceptual framework can be unifying, and seemingly risks placing them into the *relativistic equalizing* interpretation of the elephant metaphor.

What they mean here is that no particular account that one plugs into the machinery could unify the literature since the substance of the accounts change within and between various disciplines and sub-disciplines. Thus, on this kind of view, to adopt the ontology approach *just is* to allow for a diversity of content such that any account can have an incompatible alternative. I agree that for any account of addiction, it will always be possible for an incompatible view to be proposed and worked into the ontology. Again, this is part of ontology’s appeal (an additional appeal is that it would be clear where and how the views diverged). However, we are not attempting to unify a *possible* addiction literature where anyone is saying anything about it. We are trying to unify an *existing* literature where actual researchers are saying particular things about addiction. To be sure, the conversation has its problems with clarity and so forth, as I outlined above. Nonetheless, I think what West and colleagues’ point fails to capture is that most accounts of addiction are talking about the same thing.

I propose that the dispositional account of addiction makes this fact more easily recognizable. Most accounts of addiction can either easily fit into a dispositionalist framework like the one defended here, or they can be understood as already (albeit implicitly) using a dispositionalist framework, even if by another name. Most likely, the majority of cases will fall into the latter camp.

If this is right, and I aim to show that it is, I do more than simply join West, Hastings, and others in defending ontology as the key to a more integrated multi-disciplinary approach to addiction research. I also provide an account the substance of which can further unify the literature beyond what West and others had hoped. Ontology does much of the unifying, to be sure. Still, as I will go

⁹⁹ West, Christmas, et al. (2019b, p. 161), emphasis in original.

on to argue, understanding addiction as a disposition will supplement it with a further, distinct kind of unification.

Let us turn to a summary of the main lessons, and then a brief look at the chapters to come.

1.4 Lessons and Looking Ahead

My project has two main goals. First, I aim to show that addiction is dispositional in nature. This will include some proposals for the nature of the addiction disposition itself – what this disposition is *for*, *when* the disposition is liable to manifest, and so on. Second, I aim to show that my account is doubly unifying. I argue that both the content of my account (its dispositionalism) and the methodological approach (ontology) help to solve the problems of disunification pervasive in the literature.¹⁰⁰

The disunification in the literature is considerable, and more and more researchers are noticing this.¹⁰¹ However, we cannot be satisfied with pointing out that two extreme ends of the spectrum – irresistible compulsion on the brain disease models and ordinary voluntary behavior on moral or choice models – are missing something. This may well be true, but it is only the tip of the iceberg. Submerged beneath the disagreements and discussions between proponents of competing models is a crippling inability to effectively make any progress. Definitions need to be clearer and univocal. Disparate perspectives and approaches need to be interoperable, with respect to both communication and their data. We need more than just new definitions, though. The addiction literature does not simply have some unfortunate and problematic content. The real issue is with the machinery, not

¹⁰⁰ If the reader is yet unsure *how* adding one more account to the literature will do this, there is no need to worry. Nothing has been missed. I have only been setting up the project, and later chapters will go on to spell out in more detail how the methodological (Chapter 2) and substantive (Chapter 4) unification will ultimately be garnered.

¹⁰¹ This is true both generally (Alexander & Schweighofer, 1988; Goldberg, 2020; Heather, 2017b; Kalant, 1989; Sussman & Sussman, 2011; Walters & Gilbert, 2000) and with respect to the specific claim that an interdisciplinary methodology like ontology ought to be the solution (Cox et al., 2020; du Plessis, 2012, 2014b, 2018; Hastings, le Novere, et al., 2012a; Hastings & Schulz, 2012; Larsen et al., 2017; Michie et al., 2017; Michie & Johnston, 2017; West, Christmas, et al., 2019b; West, Marsden, et al., 2019b).

simply with the outputs it produces. One thing that is needed is a *method*, a way of doing things, that can by itself ensure that the resulting practice is functioning properly.

The proper methodological solution is ontology. I motivated adopting this approach as the solution to disunification, and pointed out that, while using ontology in addiction is still new, I am not the first to recognize its utility here. Both the success and the relevance of the track record ontology brings make it very hard to deny how fitting a solution it really is. As I go on to show in the next chapter, the use of ontology in other domains has quite literally solved the very problem the addiction literature suffers from. Other domains of research have been disunified, and unsurprisingly so given their complexity and the multi-disciplinary efforts involved in that research. Now, they are not so disunified. Addiction research should follow suit. I aim to help it play the right card.

Here is a short preview of what is to come. In Chapter 2, I lay out in more detail the methodology. I have said a few things about the basics of ontology to help paint a picture for the reader. We need to have more, though, in order that we more fully understand the toolbox that I am recommending we utilize going forward in addiction research.

With our method on the table, Chapter 3 begins to provide the content by filling out the substance of the dispositional account. What is the disposition like? What are its characteristic manifestations and triggering conditions? These are the kinds of questions I answer there. To be clear, I mean to say more than, for instance, that drug addiction often manifests in drug consumption that it is triggered when drugs are around. I mean to give an account of the *nature* of the addiction disposition such that it can capture the multitude of ways an addiction is manifested or triggered.

In Chapter 4, I take the content of my account and argue for its unifying power. That is, I argue that the dispositional account provides substantive unification over-and-above the methodological unification provided by implementing ontology.

Chapter 5 is comparative. With a proposed methodology and some content to plug in, I evaluate my account with respect to a number of extant views in the literature, both theoretical (accounts of what addiction is) and ontological (attempts to represent addiction in existing ontologies and classification systems). There, I introduce two more detailed sets of criteria by which these accounts and ontologies will be evaluated. As noted, they incorporate the three general desiderata introduced here. In the end, we will see why existing accounts and ontological representations are unsatisfactory, and more importantly, why the dispositional account of addiction and the implementation of ontology in the literature are preferable. Indeed, I hope to show that these are *imperative* if we are to remedy the current disunification and avoid any potential disunification going forward.

Chapter 2: Ontology and Unification in Addiction Research

2.1 Introduction and Chapter Road Map

This chapter focuses on the methodology of ontology that is interweaved into my project. It is structured as follows. This opening section provides a roadmap for the present chapter and then presents two arguments based on the discussion in the preceding chapter. The first is the overarching argument of the dissertation. The second – and the focus of this chapter – is the sub-argument for the methodological premise of that overarching argument: that ontology can help solve the problems of methodological disunification faced by the addiction literature. This, in turn, will support the methodological half of my overarching conclusion: that an ontology of addiction ought to be developed and implemented into addiction research.

Section 2.2 explains the basic toolkit of ontology in more detail. This includes a defense of the *realist* approach to ontology building.¹⁰² It also includes an introduction to some of the core principles of good ontology building underlying Basic Formal Ontology (BFO), which an ontology of addiction (or anything else) should be founded on.¹⁰³

Section 2.3 provides two examples of ontology successfully bringing unification to two domains of research. The first is the development and use of the Gene Ontology (GO) in the

¹⁰² The reader may wonder why one would need to argue for a realist approach, and what the alternative would be. I come to this below, but the short response is that an unfortunate amount of ontology work fails to respect the realist approach. This is often inadvertent, but it is by no means *always* inadvertent. The alternative approach, *conceptualism*, was born alongside ontology and so many ontologists still defend or implicitly adhere to this way of doing ontology. I will offer a simple explanation below for why this is so while arguing that this is the wrong approach, and that realism is preferable.

¹⁰³ These principles (which include *realism*) are laid out in the work of Barry Smith and colleagues and adopted by the Open Biological and Biomedical Ontology (OBO) Foundry (Arp et al., 2015; Ceusters & Smith, 2010; The OBO Foundry, 2020b). The OBO Foundry aims “to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate...[through] the development of an evolving set of principles...based on ontology models that work well...[The Basic Formal Ontology is] the upper level ontology upon which OBO Foundry ontologies are built” (The OBO Foundry, 2020b). The full lists of both BFO and OBO Foundry principles of best practice can be found in Appendices B and C.

biological sciences. The second is the development and use of the Ontology for General Medical Science (OGMS) in the medical sciences.

In Section 2.4, I recapitulate the *no difference argument* presented in Section 1.3.2 of the previous chapter, which allows us to move from the success of ontology in these other domains of research to its potential for success in addiction research.

Finally, Section 2.5 closes the chapter by recapping the main lessons learned. The central of these is that ontology will help to both:

- (i) solve the unification problems in the addiction literature; and,
- (ii) to reduce the risk of such problems resurfacing again going forward.

2.1.1 The Arguments

Chapter 1 motivated the need for the project defended here along with its basic structure. Embedded in that discussion were the beginnings of two main arguments. The first is my central, overarching argument of the entire project. The premises are to be defended throughout the dissertation, wherein I will assign one roughly premise to each chapter where it is to be more fully defended (the exception is premise (3), which gets two chapters – one filling out the dispositional account and one defending its unifying power). The overarching argument looks like this:

- (1) The addiction literature suffers from serious problems of disunification;
- (2) Ontology can help to solve the addiction literature’s problems of disunification;
- (3) The dispositionalist account of addiction is true and provides still further unification to the addiction literature;
- (4) The dispositionalist account of addiction fares better than competitor accounts, both philosophical and ontological;¹⁰⁴
- (5) If (1)-(4), then an ontology of addiction ought to be developed and implemented into addiction research and researchers ought to adopt a dispositionalist account of addiction;

¹⁰⁴ By ‘philosophical’ I just mean any account that attempts to say what addiction is, where this can include accounts from neuroscientists, psychologists, and the like. The relevant contrast is an ontological representation of addiction. The difference between these will become clearer as we move forward, and especially clear in Chapter 5.

- (6) Hence, an ontology of addiction ought to be developed and implemented into addiction research and researchers ought to adopt a dispositionalist account of addiction. [(1)-(5)]

I defended premise (1) in the first chapter, motivating the need for the account defended here.

Here is a recap. The addiction literature suffers from various problems – lack of conceptual clarity, lack of a shared vocabulary, poor definitions, theoretical problems becoming practical problems, siloing of data, incompatible methodologies, and so on. These are all problems of disunification in one form or another. Hence, *the addiction literature suffers from serious problems of disunification*.

I also *motivated* premises (2) and (3) in Chapter 1, outlining the central ideas behind what will eventually be the fuller arguments for these premises. The task of the present chapter is to move beyond motivation for premise (2) and mount a fuller defense.

Here is the basic argument for premise (2):

- (2a) Ontology provides the features that, in virtue of their absence, make a literature disunified;¹⁰⁵
- (2b) Ontology has solved problems of disunification for other domains of research which possess no relevant differences from the domain of addiction research;
- (2c) If (2a) and (2b), then ontology can help to solve the addiction literature’s problems of disunification;
- (2) Hence, ontology can help to solve the addiction literature’s problems of disunification. [(2a)-(2c)]

Combined with premise (1) from the above argument (the addiction literature is seriously disunified), premise (2c) is trivially true. In the sections that follow, I defend the remaining premises of the argument. To do so, it is essential to have a proper understanding of the fundamentals of ontology. I accomplish this task in the next section as I defend premise (2a) of the above argument.

¹⁰⁵ This need not (but can, if the reader insists) be understood causally. The *absence* of ontology – that is, the absence of implementing certain methodological principles, having a shared vocabulary, and so on – simply means that some *other* methodology is being practiced. It is thus these methods, whatever they may be, that *make* a literature disunified since certain research contexts will naturally lead to disunification unless the methods adopted are sufficient to counter this. Ontology has proven sufficient. Other methodologies have not. If *no* methodology is adopted, then it is the non-methodical (or haphazard, or random, or unprincipled, or what have you) research activities that make a literature disunified due to being insufficient for countering disunification. Using ‘in virtue of their absence’ in the premise is an easy shorthand for saying what I have just said. I try to be careful about such turns of phrase, but I also caution the reader to not let the way in which we can describe things push our metaphysics around too much.

2.2 Ontology Counteracts Disunification

This section lays out the fundamentals of ontology in more detail. There is much here, and it is worth noting why this is so. Like mathematics or logic, ontology has a stepwise nature to comprehending and using it. That is, properly understanding the fundamentals is critical for properly understanding other components, or even simple applications of ontology. Without being able to add or subtract, one's multiplication ability will suffer and doing even simple algebra will be next to impossible. Without understanding conjunction, disjunction, or the notion of a truth value, one's ability to perform even simple logical operations or to understand whether simple arguments are valid will be undermined. Likewise, for the present project we must start by nailing down the basics of ontology. Securing this foundation will ensure not only a better understanding of ontology more generally, but it will also ensure that its role and later applications in the project are not lost or misunderstood.

I start by identifying a particular problem that generates the need for ontology. I then show why this problem (and the need for ontology) is so ubiquitous, and I identify the central features of the context of this problem. With this background on the table, I turn to describing what I take to be three important goals of ontology that fall out of the foregoing problem context: carving up the world, organizing our findings, and representing reality. Through an explanation of these goals, I present and clarify the fundamental components and methods of ontology. I also explain some core principles of best practice, placing special emphasis on the principle of *realism*. Support for the claim that ontology facilitates unification (premise (2a)) is provided throughout Section 2.2.

2.2.1 A Contextual Primer: The Need for Ontology Is Ubiquitous and Unsurprising

Conceptual and definitional confusions and misunderstandings occur all the time, in virtually every context. When we communicate with others, the importance of *knowing what each other means* to

the success of the conversation cannot be overstated. This may even seem obvious. Indeed, most people are aware (or can easily be made aware) that they suffer quite often from very simple misunderstandings of meaning in their communication with others. In short, communication is really, really hard, and no one should find this very surprising.

However, when communicating, this fact is likely the last thing on anyone's mind. Most people typically take themselves to understand each other, especially when disagreeing, and proceed in conversation with this assumption taken on board. So, despite our awareness that communication is exceedingly difficult, it is common practice for interlocutors to assume that they understand each other's meaning by simply hearing the words that are uttered.¹⁰⁶ So it is unsurprising that communication often goes awry (and often does so unnoticed).

Because of this, we should not find it surprising when this happens in domains of academic or scientific research. After all, these are institutions filled with researchers from diverse backgrounds that can vary in terms of their training, research, methodology, or vocabulary. They come from labs, departments, institutions, disciplines, or research paradigms that have their own aims, terms, and background assumptions. This is all fuel for the fire.

2.2.1.1 The Need for Intervention

So, what is the lesson? Researchers in any domain are likely to fail to successfully communicate, probably quite often, and so methods for avoiding, minimizing, and remedying such problems are invaluable. Perhaps lab mates will do better than researchers working in different fields. One would hope; but we need more than lab mates to communicate well with one another. It is quite hard to find a topic of research that is completely unrelated to other disciplines or areas of study, and thus effective

¹⁰⁶ As a quick exercise, try to recall the last time you were in (or overheard) an argument during which someone stopped the conversation and said, "I'm sorry, we may be misunderstanding each other, and I want to make sure we're on the same page and not talking past one another. How are you using that term? Here's what I mean by it..." Laughable, right?

interdisciplinary communication will be needed. There is also the problem of getting researchers in the *same* discipline, even working on the same topic, to be entirely on the same page with their terms, assumptions, and overall ability to communicate. If research endeavors are going to provide the clearest and most comprehensive understanding of their target phenomena, then successful intra- and interdisciplinary dialogue is paramount. Just as with ordinary communication, we should expect that this is not going to be easy without careful and pointed intervention.

Communication is hard enough without adding to this the need to gather data (with all of one's background assumptions), organize and label it (with all of one's background assumptions), and then transfer it into a computational form so that it can be stored and managed (searched, merged, analyzed, and so on), often with the use of computers. The difficulties accumulate quickly, and the core communication problems from above are mirrored in the context of data management. Though, data management – in particular, making it discoverable, interoperable, and so forth – is part of what intra- and interdisciplinary dialogue amount to. Thus, proper data management is as important to the success of a research domain as is successful communication.

The reader may recognize this as the problem of disunification, and perhaps further recall that ontology is the preferred solution. It is important to recognize how unsurprising we should find this problem (though this by no means makes the problem trivial). It is also important to be able to understand just what the problem is so that, as we learn about ontology and what it does, we can more easily see why it is the most fitting solution. Ontology *is* the pointed intervention needed to remedy and deter miscommunication.¹⁰⁷

¹⁰⁷ Moreover, to build on the example from the previous footnote, introducing ontology into a domain of research is quite like pausing the conversation and asking everyone to get on the same page.

2.2.1.2 Contextual Features of the Need for Ontology: Science and Heterogeneity

To understand ontology, it will help to have an idea of what ontologies are supposed to do. Below I focus on three goals of ontology building that most clearly illuminate the way in which ontology counteracts disunification. These goals fall out of the background context that motivates doing ontology in the first place. Thus, let us first understand two main features of this context.

First, *an attempt to understand the world (science, broadly speaking) calls for ontology*.¹⁰⁸ The context is not *mere* disunification; it is disunification in the context of a broadly scientific endeavor to understand the world. The paradigm cases of ontology's success have been the implementation of ontology into scientific disciplines like biology and medicine. But we need not think of this as the requirement that ontology be applied only to scientific investigations as we commonly think of them, wherein lab-coated researchers look into microscopes or transfer chemical substances between beakers. The relevant feature of the biological and medical sciences is their shared attempt to understand some portion of reality. In this way, any effort to understand the world, even if only as a means to doing something else such as navigating or manipulating it, should be seen as exhibiting this feature as well. Addiction research clearly has this feature.

Second, *the existence of large, heterogenous sets of data calls for ontology*. Given the relevance of investigating the world, the other important feature of the background context is the immense amount of data collected. As we do science (even in the broad sense I have in mind), we inevitably build up (usually very large) repositories of data. In a given domain, we might think of all the data as our collective *information base* about that domain – what our current best sciences tell us about various

¹⁰⁸ I see science and philosophy as falling on a spectrum of ways to investigate and understand the world, using a mix of empirical and conceptual tools to do so. To say that ontology is called for, then, is not to say that philosophy is not. Doing ontology involves doing some philosophy. Much of the time, doing philosophy also involves doing some ontology, as when an author lays out their definitions of terms at the beginning of their paper: “By ‘free will’ I mean ‘the ability at t to do A or to do B , where A and B are mutually exclusive. By ‘ability’ I mean ‘a disposition’. Dispositions are...” Throughout my project, I am often doing philosophy and ontology at the same time. Both are central to doing good science.

domains. It includes not only raw data from, for instance, experiments or observations, but also theories that are built and tested on the basis of the existing data we have. It even includes the research articles within the domain that discuss the data and theories surrounding the relevant phenomena.

But the information we collect must also be stored somehow. The dominant method has historically been remembering it or writing it down. However, we are now in the era of *big data*, and thanks to current technology this has seen the ability of researchers to collect, store, and manipulate massive amounts of data using computers.¹⁰⁹ This vastly increases the size that datasets can become. Yet, it will inevitably be the case that the set of all data for some domain will be heterogenous. This is because, at the very least, researchers will use different names and definitions when labeling, annotating, and managing their data. Add to this that not everyone will carve up the world in the same way, and the heterogeneity of the data worsens.

In sum, the background context for ontology involves at least two things: doing science, broadly speaking, and managing incredibly large amounts of heterogenous data (often through the use of computers). Addition research clearly satisfies both criteria.

Given this context, there are three main goals of ontology that are of interest here, each of which corresponds in some way to the background context of investigating and collecting data about the world so as to better understand it. In the next three subsections, I explain each goal. Along the way we will become acquainted with the central components of ontology. Most importantly, we will see why ontology provides the tools to counteract disunification.

2.2.2 The First Goal of Ontology: Carving Up the World at Its Joints

The first goal of ontology is *to help us carve up the world*. Ontology (in the traditional philosophical sense) is the study of *being* or *existence* or *what there is*. It asks what kinds of entities there are, how they

¹⁰⁹ See Wang (2017) for discussion of the era of big data.

are related, and what the nature of those entities and relationships are. An *entity* is anything that exists.¹¹⁰ When we try to understand the world, the first thing we do is to categorize things. If asked to simply observe, an appropriate reaction would be, “Observe what?” We always investigate *something*, and scientific disciplines exemplify this point perfectly. For instance, a biologist starts by classifying things as either *biological* or *non-biological*. Biologists study cells, but not the orbits of planets around stars; they study metabolic processes, but not weather patterns. Any time we investigate the world, we categorize things according to shared features in just this way (for instance, animate vs. inanimate or harmful vs. unharmed). Cells and metabolic processes share features that make them each biological, whereas orbits and weather patterns lack those features. Furthermore, this initial carving will involve approaching the world from a particular perspective or level of granularity.¹¹¹ Biologists and astronomers study different phenomena partly because they are investigating different levels of reality.

2.2.2.1 Entities and Relations

Once a domain of study is carved out, we then classify the various entities within it (again, based on shared features) and identify (and classify) the relations between them. For instance, the biologist classifies entities into categories (or types) such as **ORGANISM**, **CELL**, **EUKARYOTIC CELL**, **CELLULAR COMPONENT**, and **HOMEOSTASIS**. They also identify types of relations between entities, such as *binds_with*, *is_parasitic_on*, *has_cellular_part*, and *is_sub-type_of*. In addition, there is a more basic level at which we can carve up the world, wherein we can identify more (indeed, the most) basic or general kinds of entities and relations that cut across any domain of interest.

To see this, consider the distinction we have already made between the types of things in the world (entities) and how they can be related to one another (relations). These are the most basic

¹¹⁰ I follow the BFO definition of ‘entity’ here, which is the top-most (that is, most general) term in BFO and thus any ontology that uses BFO as its foundation (Arp et al., 2015, p. 2).

¹¹¹ For discussion of *levels of granularity* and *partitioning theory* in ontology, see Arp et al. (2015) and Bittner & Smith (2003), respectively.

referents of an ontology.¹¹² Consider first the different types of entities. The world is full of different kinds of things – material objects, events, spatial regions, processes, qualities. It may be common for people to think of *entities* or *things* as falling into only the first category of **MATERIAL OBJECT**. However, recall that for our purposes an entity is anything that exists. Thus, since there are entities and the relations they stand in, events, qualities, spatial regions and so on count as entities, since they are not relations.¹¹³ Let us consider a couple simple examples before getting into more detail.

The chair I am sitting on now is an *instance_of* **CHAIR**, which is a kind of entity. Additionally, *my chair's brown color* and *Bob sitting in his chair from 8-10am today* are also instances of entity types, namely, **BROWN** (a kind of color quality) and **SITTING PROCESS** (a kind of process). My chair's brownness is a *dependent* entity in the sense that it requires the existence of my chair in order to exist itself – without my chair, there would be no *my chair's brown color*.¹¹⁴ Despite its dependence, it is an entity nonetheless (after all, my chair's brown color does exist). My chair stands in the *bearer_of* relation to its brown color; it *bears* an instance of the color quality **BROWN**. The brownness is part of this relationship, too. Thus, we can describe the relationship in the other direction, noting that my chair's brownness stands in the *inheres_in* relation to my chair.¹¹⁵

¹¹² Recall that an ontology is a vocabulary, and so its components are terms. Recall further that a realist ontology is one whose terms are intended to represent reality. Moreover, ontologies represent entity and relation *types* (or *classes*) as opposed to particular *instances* of those types. Note, too, that the following discussion throughout Section 2 – what kinds of entities there are, how they relate to one another, and so on – should really be seen as mostly stipulative for our purposes. I am basically just laying out BFO since this will serve as the foundation for the ontological aspect of my project. To go back to the math and logic analogy, this is somewhat like laying out the axioms of the system. BFO does take on certain metaphysical assumptions. However, only some of these are relevant to my account of addiction, and I make this more explicit in Chapter 3. Importantly, for the purpose of doing ontology (and arguing for the implementation of ontology into the addiction literature), BFO has a decades-long history of successful ontology work (including with respect to facilitating unification in various research domains). Hence, some debates – whether processes are four-dimensional, what *inherence* or *instantiation* is, and so on – are just going to be left to the side, and the track record of BFO is going to have to speak for itself. The metaphysical elements that do matter to my account, such as the nature of a disposition, I defend (and slightly add to) in Chapter 3. Otherwise, I am simply presenting BFO (i) as a way of explaining what ontology is and does, and (ii) because an ontology of addiction ought to conform to BFO.

¹¹³ Of course, relations also exist, so one might wonder whether **RELATION** should be a *subtype_of* **ENTITY**. Perhaps this is right. As far as I know, this question has been raised to Barry Smith (BFO developer) before but not much discussion or headway was had or made (personal correspondence). However, I cannot see how this would matter to my arguments.

¹¹⁴ In contrast, my chair could still exist without *my chair's brown color*. I might paint it blue, for instance.

¹¹⁵ These relationships complement one another: **A** *bearer_of* **B** just in case **B** *inheres_in* **A**.

Consider next the process of *Bob sitting in his chair from 8-10am today*. This event is also *dependent* in the sense that its existence depends on the existence of the entities that participate in it – without me and my chair, there would be no *Bob sitting in his chair from 8-10am today*. But again, its dependence does not keep it from being an entity. The chair and I stand in the *participates_in* relation to this particular sitting process, and in doing so the sitting process stands in the complementary *has_participant* relation to the chair and me. So it goes for all processes, which depend for their existence on the one or more entities that participate in them. Processes also occur at some location in space and unfold for some duration. Thus, we can see that processes are also entities, not relations.

2.2.2.2 Continuants, Occurrents, and the Sub-Type Relation

These are some of the basic ingredients for building an ontology. Our first distinction between *entities* and *relations* provides us with the two fundamental components of an ontology. The next step is rather intuitive: distinguish between different types of entities and different types of relations (and we saw some examples just above). In BFO, **ENTITY** has two subtypes: **CONTINUANT** and **OCCURRENT**.¹¹⁶ A *continuant* is understood as an entity that continues through time, wholly present at each moment. Chairs are continuants, and you and I are also continuants.¹¹⁷ Continuants *persist* through time and change. I am identical to my twenty-year-old self, despite having undergone numerous (identity-preserving) changes.

An *occurrent*, on the other hand, is understood as an entity that unfolds over time, and thus has a *duration*. We can think of occurrents as being smeared across time, so-to-speak. They are *not* wholly present at any moment, but instead *occur* over time as they are *happening*, where only a part of the

¹¹⁶ Arp et al. (2015, Chs. 5-6).

¹¹⁷ Proponents of *four-dimensionalism*, a particular view of personal identity and persistence, disagree. I set this disagreement aside for two reasons. First, it is standard practice in ontology to classify people (organisms) as continuants, and I do not intend to disrupt this long-standing practice. Second, even if four-dimensionalism were true and our ontologies are meant to accurately represent the world, this would only require a modification of the ontology. It would be of no consequence to the overall coherence or usefulness of ontology and its methods, nor to any of my arguments within this dissertation.

occurrent entity is present at any given moment of its duration. Processes are probably the most familiar kind of occurrent. Consider a baseball game. Its participants, such as the players and the equipment (balls, bats, bases, mitts), persist through the game, but the game itself is extended over time. For instance, the first batter standing at home plate awaiting the first pitch is only a part of the game. The batter, the pitcher, and the bat and ball are continuants, and so are wholly present in that moment. The baseball game is not. It takes time to unfold, and hence is an *instance_of* **OCCURRENT**.¹¹⁸ Other occurrents include behaviors, histories, and lives.

We can now see more clearly why **COLOR** is a *subtype_of* **CONTINUANT**, since its instances *persist* through time rather than *unfolding* over time. As was mentioned, qualities like colors are distinguished from other continuants like chairs and people in that they are *dependent* on other entities for their existence. Thus, we come to a second distinction between types of entities: that between **INDEPENDENT CONTINUANT** and **DEPENDENT CONTINUANT**. Further distinctions can be made, but let us turn to distinguishing types of relations, and in particular, identifying a relation that is critical for doing ontology: the *subtype_of* (or *is_a*) relation.¹¹⁹

This most basic type of relation can be gleaned from a consideration of the categorization process we just engaged in. We started with the most general type, **ENTITY**, and then identified its subtypes, **CONTINUANT** and **OCCURRENT**. In this way, we have begun constructing a hierarchy of type-subtype relations. We then distinguished two further kinds of continuants, and we could continue classifying types of entities in this way. For instance, we might note that **QUALITY is_a DEPENDENT CONTINUANT** and that **MATERIAL OBJECT is_a INDEPENDENT CONTINUANT**. What this division process reveals, though, is that the central relation in BFO is the *is_a* relation, as this is what gets us off the ground when we begin carving up the world.

¹¹⁸ More specifically, an *instance_of* **BASEBALL GAME**, which is a *subtype_of* **PROCESS**, which is a *subtype_of* **OCCURRENT**.

¹¹⁹ An example of another distinction is between *specifically* and *generally* dependent continuants. The former includes qualities, dispositions, and roles, while the latter includes information. I focus on dispositions in Chapter 3.

2.2.2.3 An Upper-Level Ontology and More Relations

Putting all of this together, we can start to represent these most basic components and distinctions. Indeed, this is precisely what BFO does as an upper-level (domain-neutral) ontology. As discussed, we can make distinctions at domain-specific levels like biology and medicine: **EUKARYOTIC CELL** *is_a* **CELL**; **METABOLIC PROCESS** *is_a* **BIOLOGICAL PROCESS**; and so on. Correspondingly, there are domain-specific ontologies representing different types of entities and relations at the level of specific domains (we will see two examples below). However, BFO is domain-neutral because its task is to represent the most basic types of entities and relations in reality that exist across all (or most) domain of study. Such entity and relation types include those we have been discussing, and Figure 1 below depicts all of the entity types represented by BFO and their corresponding *is_a* relations:

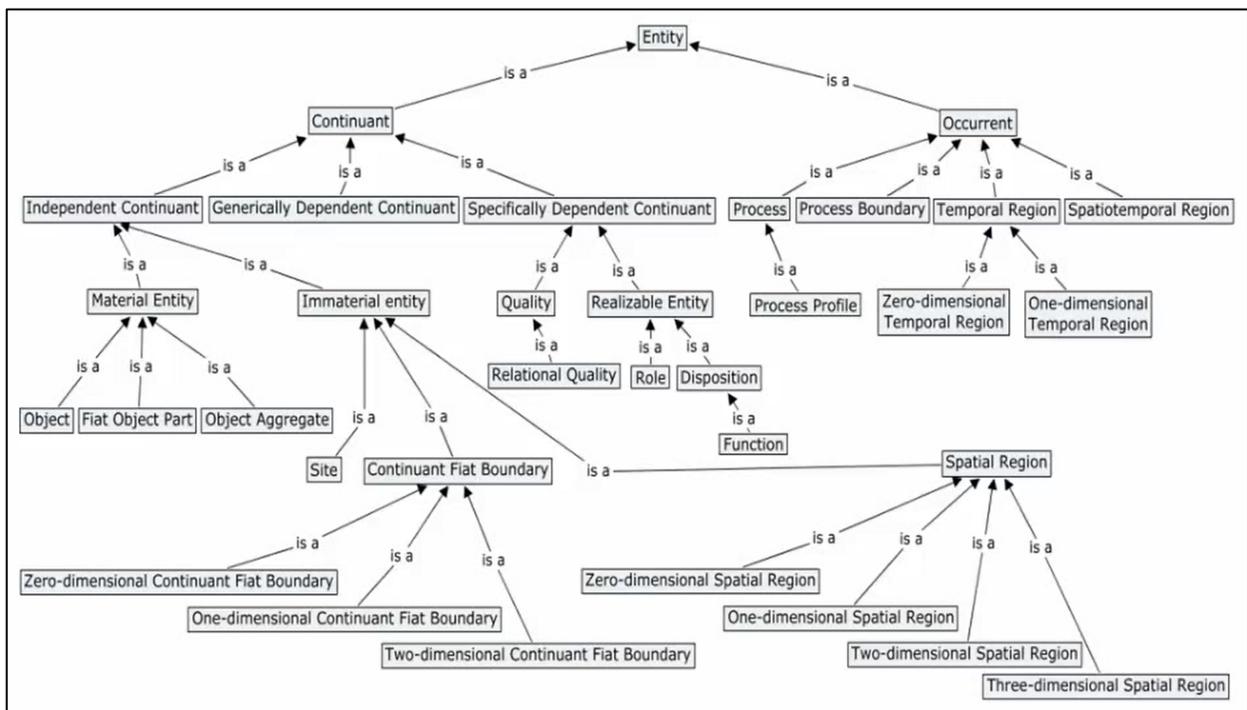


Figure 1: Entity Types in BFO¹²⁰

More relations exist in BFO besides the backbone *is_a* relation. For instance, occurrents (which unfold over time) have parts that correspond to and are present at the intermediate moments

¹²⁰ This diagram is taken from Barry Smith's presentation at <https://www.youtube.com/watch?v=bGPVCkuKT04>.

and segments of their duration. For instance, *Justin Turner hitting a home run in the first inning* was a part of the entire baseball game that was *Game 4 of the 2020 World Series*, and *the ball flying over Kevin Kiermaier's head into the center field stands* was a part of both. In BFO, processes (a kind of **OCCURRENT**) that are embedded within other processes are called 'process parts', and the corresponding relation is the *has_process_part* relation. Many continuants also have parts, such as chairs and people, and so will stand in the *has_continuant_part* relation to their respective parts (such as legs, arms, and backs). Consider Figure 2 below that illustrates the use of both the *is_a* and *part_of* relations, focusing on an organ (the pleural sac) and its related anatomical entities.

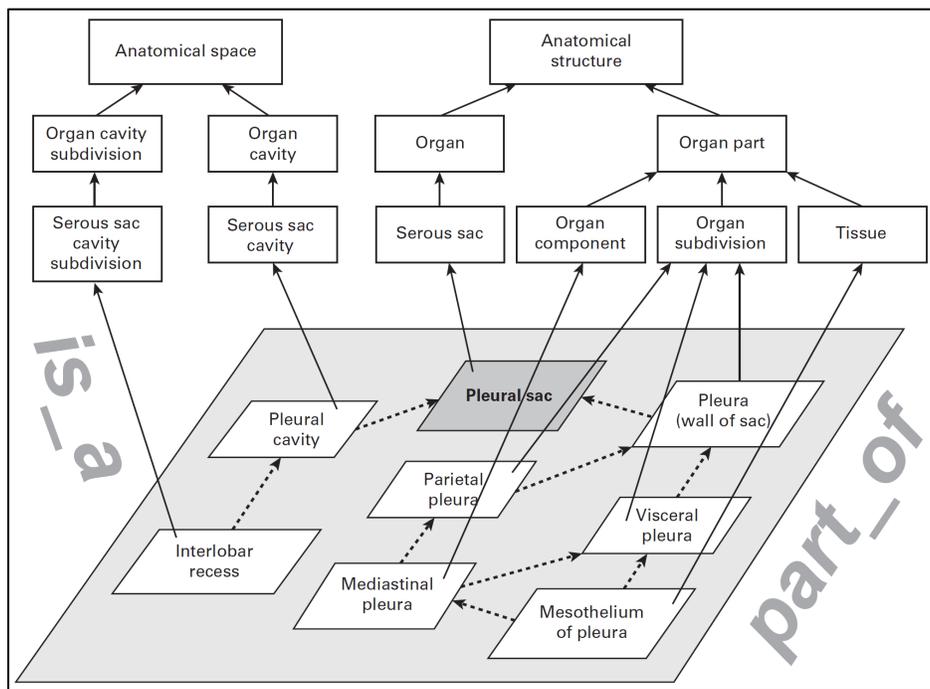


Figure 2: Representation of the *is_a* and *part_of* Relations¹²¹

Examples of other relations in BFO are the *bearer_of*, *inheres_in*, *has_participant*, *located_in*, and *instance_of* relations.¹²² Consider Figure 3 below which represents the process subtype **CLINICAL REVIEW TRIAL** and the entities and relations involved in it. This diagram helps to illustrate how we can represent many entity and relationship types at once.

¹²¹ This diagram is taken from Arp et al. (2015, p. 34).

¹²² See Buffalo Developers Group (2020).

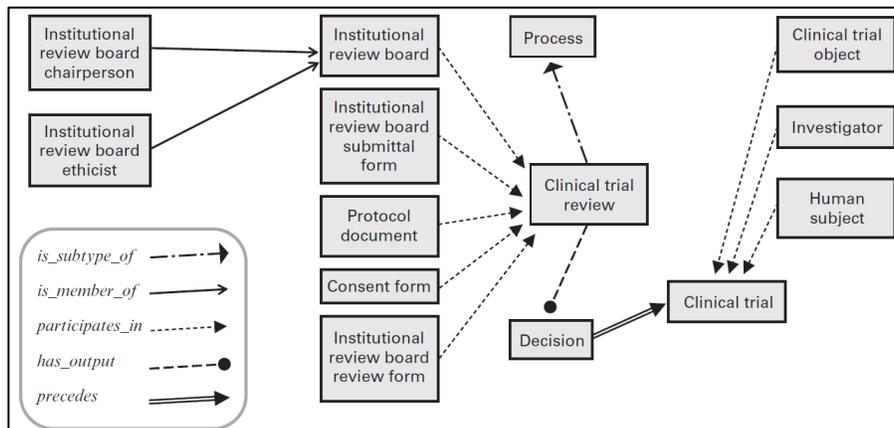


Figure 3: Example Diagram Representing Multiple Types of Entities and Relations¹²³

As is likely clear by now, classifying entities into types and sub-types (via the *is_a* relation), and identifying the relations they stand in, can quickly become quite intricate. This is unsurprising, since the world is quite an intricate place. Figure 4 below is an example of an ontology diagram representing a more intricate carving up of a (relatively small) portion of reality related to nanoparticles.

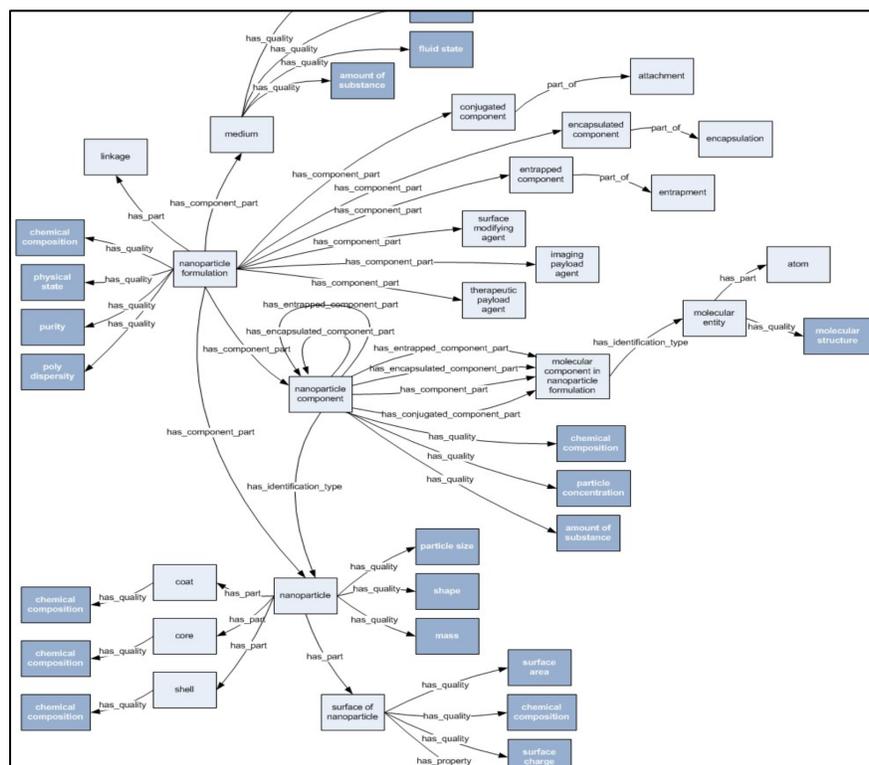


Figure 4: Example of a Complex Ontology Diagram¹²⁴

¹²³ This diagram is taken from Arp et al. (2015, p. 35).

¹²⁴ This diagram is taken from the NanoParticle Ontology homepage: <http://www.nano-ontology.org/>.

2.2.2.4 Taxonomies: The Backbone of Ontologies

The point here is simply that when carving up the world, our classifications can quickly become complex. This is not necessarily a problem – again, the world is in fact complex – but is instead all the more reason to have a careful, systematic way of doing the carving. In ontology, the fundamental relation in carving up the world is the *is_a* relation. This means that the backbone to any ontology is a hierarchical classification of entity types classified into type-subtype relations based on common essential features. Some of the oldest and most familiar examples of ontologies are *taxonomies*, and the most familiar of these is probably the taxonomy of organisms into species, genus, and so on. Figure 5 below is an example of a taxonomy. It depicts the *five-kingdom classification* for living organisms:

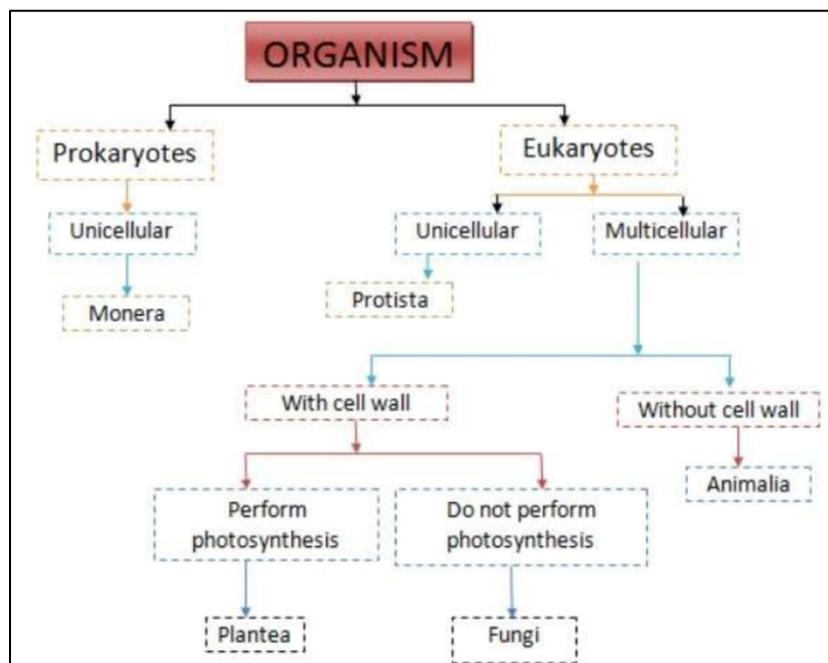


Figure 5: The Five Kingdom Classification System of Living Organisms¹²⁵

Like in the diagram from BFO, the boxes (called ‘nodes’) represent types of entities and the lines (called ‘edges’) represent the relations that hold between them (such as *has_subtype*). Consider the

¹²⁵ This diagram is taken from the slides at: <https://www.slideshare.net/saraswatimedidi/classification-of-organisms-56586269>. The arrows in this diagram do not strictly follow proper ontology principles. For instance, the nodes under ‘Eukaryotes’ should really say ‘Unicellular Eukaryote’ and ‘Multicellular Eukaryote’. Still, this will do for our purposes. See the Encyclopedia Britannica entry on this system for more detail and the history of its development: <https://www.britannica.com/science/taxonomy/Current-systems-of-classification>.

type (or class) **FUNGUS**.¹²⁶ From the above taxonomy, we know that **FUNGUS** *is_a* **EUKARYOTE**, which *is_a* **ORGANISM**. We also know that we could extend **FUNGUS** down to include **MUSHROOM** as a subtype, since mushrooms are fungi. From this, we can conclude that **MUSHROOM** *is_a* **ORGANISM**, since the *is_a* relation is transitive.¹²⁷ Included in this transitivity is the downward inheritance of essential properties, so we can also conclude that mushrooms are *multicellular*, that they contain *cell walls*, and that they *do not perform photosynthesis*.

But a question arises as to how we further distinguish types of entities. As the taxonomy illustrates, we do not merely rely on the basic *is_a* relation between a subtype (called a ‘child’) and the entity of which it is a subtype (called a ‘parent’). This is because entity types often have multiple children, each of which stand in the *is_a* relation to the parent (two children of the same parent class are called ‘siblings’). So, while it is true that **FUNGUS** *is_a* **EUKARYOTE**, it is also true that **PLANT** *is_a* **EUKARYOTE**. The same goes for **ANIMAL** and **PROTISTA**. Hence, we need more than the *is_a* relation between child and parent to distinguish siblings of the same parent. Figure 5 above illustrates the beginnings of this. For instance, **ANIMAL** and **PROTISTA** are distinguished on the basis of their cellular makeup, such that **ANIMAL** *is_a* **EUKARYOTE** *that is unicellular* and **PROTISTA** *is_a* **EUKARYOTE** *that is multicellular*. This is to utilize Aristotelian definitions, and it is central to ontology building.¹²⁸

2.2.2.5 After the *is_a*: Aristotelian Definitions

An Aristotelian definition uses the form **A** *is_a* **B** *that Cs*, where ‘A’ refers to the entity type being defined, ‘B’ refers to the parent type, and ‘Cs’ refers to those features which differentiate instances of **A** from instances of **B**’s other children (sometimes referred to as ‘differentia’). This is in

¹²⁶ I use singular terms for labeling types, since this is a recommended principle of best practice, which we will come to below. I also use the English words when discussing the nodes below for simplicity, such as ‘plant’ instead of ‘plantea’.

¹²⁷ In other words, if **A** *is_a* **B**, and **B** *is_a* **C**, then **A** *is_a* **C**. Moreover, **A** will share any features that **B** and **C** have *essentially* – that is, features that *make them* a **B** and a **C**, respectively. This is even spelled out somewhat in the taxonomy in Figure 5, though this is not quite how an ontology diagram should be drawn. Again, though, it will do for our purposes.

¹²⁸ See Arp et al. (2015, Ch. 4).

line with what we have already been doing, which is carving up the world by classifying entities into types, subtypes, sub-subtypes, and so on, based on shared, essential features. Consider the primary distinction from above between **CONTINUANT** and **OCCURRENT**. The class **ENTITY** is their parent (indeed, it is the top-most parent for any class in a BFO-conformant ontology). Recall that what distinguishes these two subtypes of **ENTITY** – the *differentia* – is that continuants persist through time and occurrents unfold over time. Thus, using Aristotelian form (and drawing on BFO¹²⁹), we can define these entities as:

CONTINUANT (A) *is_a* **ENTITY** (B) *that persists, endures, or continues to exist through time while maintaining its identity* (Cs);

OCCURRENT (A) *is_a* **ENTITY** (B) *that unfolds itself in time or it is the instantaneous boundary of such an entity (for example a beginning or an ending) or it is a temporal or spatiotemporal region which such an entity occupies* (Cs).

In this way, the Aristotelian definition highlights two important features. First, it classifies the type of entity being defined (A) as a child of some parent type (B), thereby demarcating it from other types or classes of entities (non-Bs). Recalling the inheritance of essential properties that the *is_a* relation entails, this tells us that A has any property, or stands in any relation, that B has or stands in essentially (or by definition). This is part and parcel with the transitivity of the *is_a* relation. As in Figure 5 above, if **FUNGUS** *is_a* **MULTICELLULAR EUKARYOTE** *that has cell walls*, then any child of **FUNGUS**, such as **MUSHROOM**, will be such that all of its instances also have cell walls.¹³⁰

Hence, the second feature the Aristotelian form highlights is that the entity type being defined (A) is demarcated from any siblings through the identification of *differentia* (Cs). **A** is not simply a **B**; **A** is a **B** *that Cs*. Thus, as in Figure 5, **FUNGUS** is not simply a kind of **MULTICELLULAR EUKARYOTE**

¹²⁹ For English language definitions and elucidations of BFO's entities and relations, see Buffalo Developers Group (2020).

¹³⁰ Note that we do not have to build this into the definition of 'mushroom', but it instead comes for free. That is, we do not need to add 'that has cell walls' to the *differentia* component of the definition of 'mushroom'. The part of the definition that states '**MUSHROOM is_a FUNGUS**' already entails this, since fungi have cell walls by definition. We would only need to add any *differentia* that are essential to distinguishing **MUSHROOM** from any other children of **FUNGUS**.

that has cell walls. This is true, but it does not yet distinguish **FUNGUS** from **PLANT**. Fungi do not perform photosynthesis, so **FUNGUS** *is_a* **MULTICELLULAR EUKARYOTE** *that has cell walls and does not perform photosynthesis.* The point here is that Aristotelian definitions help us to carve up the world into types and subtypes based on common, essential features.¹³¹

2.2.2.6 Universals and Defined Classes

So, the nodes in our diagrams (and the terms in realist ontologies) represent entity types and subtypes, but what exactly does this mean? According to BFO, the types represented by the nodes are either universals or classes defined in terms of universals. *Universals* are roughly understood as repeatable entities that are *instantiated* by particular (non-repeatable) in-the-world entities. For instance, **OBJECT**, **PROCESS**, and **QUALITY** would be universals in this sense. Particular objects, such as the red ball in Allie’s closet and the blue pen here on my desk, each instantiate the universal **OBJECT**. Their respective colors also instantiate the universal **QUALITY**. It is in this sense that universals are said to be repeatable. If **HUMAN BEING** is a universal, it would work in the same way. You and I are particular human beings, and so we are each instances of **HUMAN BEING**. Instantiating universals (or types) can roughly be understood as *bearing the properties and standing in the relations in virtue of which something counts as an instance of that type.*¹³²

Classes defined in terms of universals – *defined classes* – are somewhat similar in that their members consist of collections of instances that share certain common features that, in virtue of this fact, make each particular count as a member of that defined class. The idea here is roughly that we can group things according to features that we care about, even if they do not correspond to what we might call a ‘real’ or ‘natural’ grouping. This is why they are understood as classes that get defined *in*

¹³¹ In other words, **A**’s being a subtype of the parent class to which it belongs (**B**) and having the features that differentiate it from its siblings (*that Cs*) must be essential to what it is to be that thing (Arp et al., 2015). However, this is for *universals* as opposed to *defined classes*. We come to this distinction in the following section.

¹³² This is not a definition. It is just an attempt to elucidate the idea.

terms of universals. The class **RED OBJECT WITH STICKERS ON IT** is almost certainly no more than a made-up class defined according to some interest or goal we might have (or perhaps just randomly). Each member of this class becomes so by being an *instance_of* **OBJECT**, being *bearer_of* some instance of the quality **RED**, and (something like) being an appropriate *participant_in* some instance of a **STICKER PLACING PROCESS**. Other examples of defined classes might be **SMOKER**, **WRISTWATCH**, **PUBLISHED ARTICLE**, and so on.

We categorize things on the basis of shared features. However, not all categorizations correspond to what might be thought of as *genuine* differences in the world. In BFO, defined classes differ from universals in that the latter are thought to capture genuine types in reality. They carve nature at its joints, as it were, whereas defined classes capture any other groupings we might care to make, such as for some specific practical purpose. There is probably not a universal (in this BFO sense) corresponding to **SMOKER**. It is more likely a pragmatic way of classifying people according to certain shared features or behaviors that are of particular interest to us – for instance, because we might care about the correlation between smoking behaviors and cancer. We can, of course, still make distinctions like this and classify types of entities accordingly.¹³³

2.2.2.7 Whales Are Not Fish: Providing a Shared Vocabulary

At this point, we should notice that carving up the world in this way already begins to provide us with a shared vocabulary. Each node in the diagram is defined in terms of the parent class that it falls under, along with some *differentia* that distinguish it from its sibling classes. This is just the Aristotelian form (**A is_a B that C**). The simple taxonomic structure serves as the backbone of the ontology, providing the *is_a* relations between each entity type within it. The definitions given then

¹³³ It is worth noting again that I am simply laying out the BFO framework so that the reader has a solid understanding of the fundamentals of BFO-style ontology. This is relevant to the practical component of my project, namely, my argument that an ontology of addiction ought to be developed, implemented into the addiction literature, and conformant with BFO.

further specify the relevant features (the *differentia*) on the basis of which the classification is made. So, we can see how merely constructing a hierarchical backbone taxonomy with Aristotelian definitions simultaneously serves as the process for building a simple and clear vocabulary that can easily be understood by anyone who uses the ontology. It provides a shareable vocabulary.

The taxonomic *is_a* structure and the corresponding definitions are each equally important. To see this, consider Figure 6 below, which classifies animals into the following types: **MAMMAL**, **FISH**, and **REPTILE**.

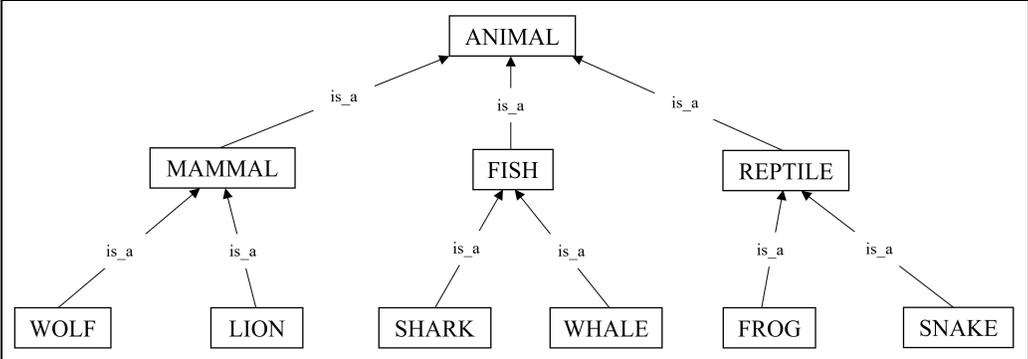


Figure 6: Example of a Possible Ontology Classifying Whales as Fish

Of course, one might first notice that whales are not *actually* fish, and so conclude that the taxonomic structure of this ontology is incorrect. However, this conclusion is hasty without first knowing what the definitions are at the relevant nodes. For instance, suppose ‘fish’ is defined in this ontology as *animal that lives in water*, or in Aristotelian form: **FISH** *is_a* **ANIMAL** *that lives in water*. Perhaps the creator of the ontology was classifying animals according to their environment, and further defined ‘mammal’ and ‘reptile’ as *animal that lives on land* and *an animal that lives on both land and water*, respectively. If so, the ontology is perfectly accurate.¹³⁴

Admittedly, it may sound odd to say, “A whale is a kind of fish.” But this is only because we have something other than *animal that lives in water* in mind as the meaning of ‘fish’ – perhaps something

¹³⁴ What is more, it is likely that you did not think twice about **FROG** and **SNAKE** being subtypes of **REPTILE**, even though ‘a frog is a reptile’ has a completely different meaning in this ontology from what you likely had in mind. This further illustrates the importance of knowing the definitions as well as the taxonomic *is_a* structure.

like *animal that lives in water and breathes through gills* or some such thing, which would make the statement obviously false. The point is that it is crucial to know *both* the taxonomic *is_a* structure *and* the definitions of the terms. Once we do, we can easily see that our initial reaction does not actually express a disagreement with the ontology – we simply mean something else by ‘fish’. We may not *like* the use of ‘fish’ for labeling such a classification; we may also think ‘reptile’ is the wrong label and prefer ‘amphibian’ to be used to name types of animals that live on both land and in water. Nonetheless, it is incredibly useful to know that the disagreement with the ontology is not about the world, but only about what certain groupings of entities are rightly called.

Ontologies provide a shared vocabulary. It starts with the backbone *is_a* hierarchy and includes Aristotelian definitions that specify further *differentia*. In addition, ontologies typically include further information alongside the definitions, such as synonyms, elucidations, examples, and comments describing, for instance, common but non-necessary features of the class being defined.¹³⁵ Ontologies also specify additional relations that exist between types of entities beyond the *is_a* relation. These can help to clarify the meaning of the terms, as well as what the ontology is attempting to represent and how. The vocabulary is *shareable* in that anyone who uses the ontology would be using the same definitions for their terms. The definitions are also formed as simply as possible and made entirely explicit, which helps to mitigate misunderstanding and to make mere verbal disagreements easier to detect (such as the one about whales being fish).

2.2.2.8 Goal One and Unification

Let us take stock. The first goal of ontology helps us to carve up the world. A paradigm example of carving up the world is doing science, broadly construed, wherein we investigate reality in order to determine what kinds of things exist, what they are like, and how they are related to one

¹³⁵ For instance, see the Gene Ontology classes here: <https://bioportal.bioontology.org/ontologies/GO>.

another. Moreover, scientific investigation produces large sets of heterogeneous data. Thus, given the general difficulty in communicating it is unsurprising that this background context produces a widespread need for the use of ontologies. Here are some of the main points we have learned up to this point about the fundamentals of ontology:

- (i) Ontologies are used to help us carve up the world;
- (ii) The fundamental components are entity types and the relations between them;
- (iii) Entity types are classified into type-subtype relations according to shared essential features, and the two central types are **CONTINUANT** and **OCCURRENT**;
- (iv) While there are many relations, the backbone structure is a hierarchical taxonomic classification consisting of the main type of relation: the *is_a* or *is_subtype_of* relation;
- (v) Entity types are either *universals* or *classes defined in terms of universals* (“defined classes”);
- (vi) Universals and defined classes are given Aristotelian definitions: **A is_a B that C**;
- (vii) Ontologies can be (more and less) domain-neutral and domain-specific; and,
- (viii) Ontologies provide an explicit, sharable vocabulary.

Recall that a central feature of disunification in a domain is that researchers are not on the same page about their terms. When people understand and use terms differently, especially when this is not explicit, there is a risk of talking past one another. Furthermore, disunification in the addiction literature is exacerbated by many accounts having unclear, problematic, or altogether missing definitions or characterizations of terms and concepts central to the domain. This easily leads to inconsistencies and conceptual confusions. Finally, a central feature of disunification is the absence of a shared, systematic method for addressing these kinds of problems.

Understanding the first goal of ontology helps to demonstrate that ontology is the way to remedy such problems of disunification. It provides a clear and systematic way of classifying entities into types and sub-types according to shared essential features. The upper-level BFO provides a foundational classification of domain-neutral entity types from which more precise, domain-specific ontologies can be developed. The backbone *is_a* hierarchy and the Aristotelian definitions constitute a systematic method for providing clear, explicit definitions of terms. Moreover, further clarification

is provided through the use of synonyms, elucidations, examples, and comments. Finally, at least as important as the method for developing it is the shared vocabulary that is produced in constructing an ontology.

For all of these reasons, ontology helps to combat disunification in a research domain. In other words, premise (2a) – *ontology provides the features that, in virtue of their absence, make a literature disunified* – has found some support. I turn now to some additional support in the second goal of ontology.

2.2.3 The Second Goal of Ontology: Organizing Our Findings

The second goal of ontology is *to organize our findings about the world*. As noted, an investigation of the world (science, broadly construed) inevitably leads to compiling a repository of data. Observations and experimental results are recorded, and these data and the theories built around them are discussed, explained, and developed in things like works of journalism, popular science books, and academic articles and textbooks. Perhaps most importantly, though, is the fact that in our present context of big data and computer technology, more and more of the findings in a given domain of research are translated into computational forms in order to facilitate more efficient and effective data management. Hence, it is unsurprising that there is a wealth of data for any research domain, much of which is made machine-readable and stored in computers. This is especially unsurprising for highly interdisciplinary domains like addiction research.

So, what exactly is the problem? Are the researchers themselves not organizing their data as they compile, record, and translate it? What role does ontology play in helping to organize our findings? The remainder of Section 2.2.3 answers these questions.

2.2.3.1 Heterogeneous Data Calls for Ontology

The problem with organizing our findings about the world is not *merely* that there are such massive amounts of data. It is instead that so much of that data is heterogenous. This stems from the

fact that data is collected by an innumerable number of researchers, investigating the world from within various disciplines and theoretical frameworks, and using different methods for managing their data. For instance, Lidong Wang has noted that there are four types of heterogeneity of data within the context of so-called ‘Big Data Analytics’:

There are [the] following types of data heterogeneity [2]:

- *Syntactic heterogeneity* occurs when two data sources are not expressed in the same language.
- *Conceptual heterogeneity*, also known as semantic heterogeneity or logical mismatch, denotes the differences in modelling the same domain of interest.
- *Terminological heterogeneity* stands for variations in names when referring to the same entities from different data sources.
- *Semiotic heterogeneity*, also known as pragmatic heterogeneity, stands for different interpretation of entities by people.¹³⁶

Consider also the way that Yongqun He and colleagues describe this problem, which they call the ‘five V’s’, and point to ontology as the solution:

[The first bottleneck hindering data integration and the deepest understanding of a domain possible is that] the characteristic five V’s of our Big Data era lead to disintegrated and non-interoperable data and knowledge. The amount of data (volume), speed at which it is produced (velocity), range of its sources (variety), quality and accuracy (veracity), and assessment of utility (value), result in large, complex, multidimensional, and diverse datasets. Disintegrated and non-interoperable data cannot be interpreted by computers and this inhibits computer-assisted reasoning, which is the essence of artificial intelligence. Consequently, our knowledge – data and information that embodies awareness and understanding – of domains represented by various datasets is seriously hindered...A critical key to data/information/knowledge disintegration and big data analysis is ontologies...[which] robustly support data integration, sharing, reproducibility, and computer-assisted data analysis.¹³⁷

The point here is twofold. Heterogeneity is a commonly recognized problem in the context of organizing and synthesizing data, and it ultimately amounts to variability in types, formats, expressions, and interpretations of data. To see why heterogeneity can be problematic, consider a simple example.

Suppose a medical researcher at a hospital is working on a vaccine for COVID-19, and they want to collect data from hospitals around the country on things like survival rates, durations of disease courses, comorbidities, and so forth. Now suppose further that some hospitals use ‘COVID-19’ to label (or ‘tag’) their data, intending to refer to the disease itself, while others use ‘SARS-CoV-2’ to label

¹³⁶ Wang (2017, p. 8), emphasis added.

¹³⁷ He et al. (2020, p. 2).

their data, intending to refer to the virus causing the disease. In each case, they use their respective terms to tag their data representing the number of patients diagnosed with coronavirus disease. We might imagine further some hospitals using ‘CVD19’ instead of ‘COVID-19’, and any number of other variations. A simple table representing the number of cases might look like Table 1:

Institution	Coronavirus Disease Cases
Hospital A	n = ‘COVID-19’ = 1,492
Hospital B	n = ‘SARS-CoV-2’ = 928
Hospital C	n = ‘CVD19’ = 309
...	...

Table 1: Number of Cases of Coronavirus Disease at Hospitals

If this does not seem so bad yet, consider the fact that the researcher could not have compiled even this simple table without doing the following: accessing each hospital’s database; discovering which label (‘COVID-19’ or whatever it happens to be) represents the number of cases of coronavirus disease at that hospital; and entering that number into their own database under some label (such as ‘Coronavirus Disease Cases’). They could not just merge the data together because the computer would already need to know that ‘COVID-19’, ‘SARS-CoV-2’, ‘CVD19’, and the like all had the same meaning in the datasets.¹³⁸ Hence, one would have to do that digging anyways.

Add to this any number of additional complexities that heterogeneity of the data produces. For instance, perhaps not all hospitals distinguish between diagnoses and misdiagnoses because some fail to modify their intake data when a patient is discovered to have been misdiagnosed. If so, simply adding up the total number of reported cases across hospitals would be unhelpful since some misdiagnoses would be counted as genuine cases. Alternatively, suppose we needed to compare individual cases, but hospitals use different labels for patients such as random identification numbers, patients’ names, the last four digits of their social security number, or something else. Another possible

¹³⁸ The computer does not get the meaning of terms based on the English language labels we give to our data.

confound is that some patients might visit multiple hospitals or take a different type of test at the same hospital, and so might be double counted.¹³⁹

In short, heterogenous data is extremely difficult to synthesize, even with the most sophisticated software tools. A computer cannot call the hospital and ask what they mean by ‘CVD19’ in their data set, or whether they already accounted for misdiagnoses. However, if each hospital possessed a shared vocabulary of terms to tag their data, accounted for the difference between disease and diagnosis (an assertion which is about some purported instance of a disease¹⁴⁰), and used unique identifiers to refer to each kind of entity, the synthesizing of their data looks to be much, much easier and more coherent. It is in this way that heterogenous data calls for ontology.

I turn now to two ways in which ontology helps us to organize our findings about the world.

2.2.3.2 The First Type of Organization: Classification

There are two basic ways in which ontology helps to organize our findings. The first is quite simple and concerns the basic structure of the ontology itself, which we have just been discussing. This includes the *is_a* hierarchy, the Aristotelian definitions, and the specification of relations. In other words, the process of carving up the world through ontology building already begins to organize one’s findings in a systematic and useful way. This is especially so if the ontology is linked to and conformant with the domain-neutral BFO, since this helps move us beyond an understanding of the relevant types of entities and relations just in the context of some specific domain, such as biology. More specifically, a BFO-conformant ontology can help move us towards a deeper understanding of the nature of those types of entities and relations, and how they are related to others in different domains.

¹³⁹ There have been reports of some actual double counting of COVID-19 cases (Stolberg et al., 2020).

¹⁴⁰ See Limbaugh et al. (2020; 2019).

Ontology is a robust form of classification, and classification is a form of organization. Hence, the very structure of an ontology offers a way to help us organize our findings into types of entities with specified differentia that stand in particular kinds of relations to one another.

2.2.3.3 The Second Type of Organization: The Semantic Web

The second way in which ontology helps to organize our findings concerns how the ontology can be stored and used in the context of the Semantic Web.¹⁴¹ The Semantic Web is best understood in contrast to its precursor that most are more familiar with: the World Wide Web (WWW). The WWW is, roughly, a mechanism that allows information from disparate sources to be linked together by linking electronic documents. Sometimes these documents are files (Word documents, PDF files, etc.) that are uploaded to webpages, and sometimes the document is the webpage itself, which might contain images, text, and so on. The main idea is that the WWW roughly consists of a series of links between documents that contain information (like a social media profile or a journal article about Plato's *Republic*), and have a unique identifier (sometimes described as a unique 'location' on the web), which we know as a *uniform resource location* (URL). This allows ease of access to information by linking one document at a given URL to another document at a distinct URL.¹⁴²

For all of its incredibly useful features, the basic limitation of the WWW relevant to our purposes is that it links *documents* rather than the *particular pieces of information* (the data or facts) contained within the documents. A computer interprets a URL as saying something like, "This document here containing information x, y, and z." So, using a computer to query information about x, y, and z will likely return the just-mentioned document, but it will also return every other document that simply

¹⁴¹ The following discussion of the Semantic Web draws on the Cambridge Semantics introduction to the topic (<https://www.cambridgesemantics.com/blog/semantic-university/intro-semantic-web/>) and from Coursera's Massive Open Online Course (MOOC) entitled, "Web of Data" (<https://www.coursera.org/learn/web-data>).

¹⁴² We sometimes refer to this as 'following a link'. For instance, one online article might link to a news article on a related topic mentioned in the original article. Another example is navigating from a restaurant's homepage to its menu subpage.

mentions ‘x’, ‘y’, and ‘z’, even if they are *not* relevant.¹⁴³ Moreover, it will likely *not* return documents that do not mention ‘x’, ‘y’, and ‘z’, even if they *are* relevant. This is because the computer is not looking for x, y, and z *themselves*, but is instead looking for *documents* (URLs) that *mention* ‘x’, ‘y’, and ‘z’.

This feature of the WWW explains why modifications to information in one document does not spread to other documents containing that information. Imagine taking a copy of *The Stranger* and changing the name ‘Meursault’ to ‘Baudelaire’. We would not expect every other copy of *The Stranger* to now read ‘Baudelaire’ where ‘Meursault’ once was. This is how a URL works: changing the document means that any link to that document will bring up the newly modified information, but it will not change the information within other related documents; this is so even if they are all linked to the original document that was modified. This is because it is the *documents* that are linked and not the *particular pieces of information* within them. Enter the Semantic Web.

The basic idea behind the Semantic Web is that it took the concept of identifying, locating, and linking *documents* (web pages) using unique URLs, and extended it to identifying, describing, and linking *individual pieces of information*. Ultimately, this was a way to link the facts that the information represents. It does this by using unique identifiers called ‘uniform resource identifiers’ (or ‘URIs’).¹⁴⁴ A resource can be understood as anything that we can describe (and hence, give a unique identifier to), such as a person, an umbrella, a baseball game, a bank’s auto loan process, and so forth. In this way, URIs and the Semantic Web allow data and the facts they represent to be linked, rather than just electronic documents that mention them. We can also have URIs for relations since we can describe what it is for something to stand in some relation R. Hence, URIs give us a way to link different resources (people, documents, properties, events) via the relations they stand in to one another.

¹⁴³ Hence, you may not find the documents that are relevant without scrolling through hundreds of irrelevant links.

¹⁴⁴ These typically look much like the normal URLs (web addresses) we are familiar with.

The upshot is that instead of modifying and linking only *documents* that are *about* resources, we can link and modify the *resources themselves* (or at least representations of them). Changing Meursault's name to 'Baudelaire' would amount to modifying the relation between two different URIs (one representing Meursault and one the *has_name* relation) and particular strings of letters. Consider the following modification of Meursault's name to 'Baudelaire':

Original: Meursault (URI #1) *has_name* (URI #2) 'Meursault' (string #1)

Modified: Meursault (URI #1) *has_name* (URI #2) 'Baudelaire' (string #2)¹⁴⁵

We could also change the description of Meursault to reflect that he lives in Belarus instead of Algiers:

Original: Meursault (URI #1) *lives_in* (URI #3) Algiers (URI #4)

Modified: Meursault (URI #1) *lives_in* (URI #3) Belarus (URI #5)

Now, how does this help us to organize our findings about the world? To see this, consider first what the Semantic Web and the use of URIs does for ontologies. An ontology classifies and defines types of entities and relations – for instance, **A is_a B that Cs**, **A has_part D**, **A participates_in E**, and so on. Since each type of entity and relation would have a unique URI, we can think of the URIs as analogous to URLs for web pages and electronic documents. The URI is like having a web page that describes everything known about that type of entity or relation; it is somewhat like giving it a social media profile. So, when an entity type in an ontology – more specifically, a *term* in the ontology representing some entity type – is used to tag data, this links that data to a particular URI. In the ontology, this URI is linked to various other entity types (URIs) via various kinds of relations (URIs). Hence, tagging some piece of data with the original URI brings with it all of the corresponding information at that URI – that is, about that entity type – for free.

¹⁴⁵ Two things to keep in mind. First, Meursault is distinct from the string of letters 'Meursault' that we use for his name. Second, we do not need a URI to represent a string of letters since the string itself is uniquely identified by its components.

Hence, the Semantic Web can be seen as a place to put our ontologies in the same sense that electronic documents were placed and linked up on the WWW. The difference is that it links individual facts and data. One could create a webpage (with a unique URL) describing their ontology, complete with graphs, visualizations, and links to each entity and relation type utilized (a kind of sub-page) that contain their definitions, synonyms, and so forth. The ontology (more specifically, its terms, corresponding definitions, and so on) can be used by others when discussing those kinds of entities or tagging data about them.

Thus, in addition to organization through classification, ontologies help to organize our findings through use of the Semantic Web. First, by linking individual meanings (descriptions of resources) instead of whole documents, the Semantic Web provides an easy way for a structured vocabulary like an ontology to be shared. Having a shared vocabulary is a central key to unification. Second, the use of URIs ensures that the vocabulary is *shared* – that is, the meanings of the terms are preserved – and thus that verbal disputes are avoided. This is because the URI provides the meaning it does independently of the word or phrase we use to label it. As an analogy, consider that it does not matter what we call a website; so long as we use the same URL, we will get the same information. Third, through the Semantic Web and the use of URIs, we can tag our data in a way that is consistent and free of ambiguities. For instance, two institutions studying the effects of the direct stimulation of the nucleus accumbens on cocaine craving and ingestion behaviors in rats would be able to utilize the same domain ontology containing the relevant entities and relations.¹⁴⁶ This would make explicit what they mean by ‘canula’, ‘craving’, ‘ingestion’, ‘behavior’, and so on. How would we know? Because the researchers would have tagged each piece of their data with either the same or different URIs.

Thus, ontology helps to organize our findings about the world in two ways:

¹⁴⁶ Likely, they would use a number of ontologies, but the idea is that they could tag their data with the same terms from the same ontologies. More importantly, even if they failed to do this – such as one describing the rat as ‘addicted’ and the other as ‘dependent’ – they would be able to recognize if and where they disagreed.

- (i) the very practice of building ontologies through classification into types of entities and relations between them; and,
- (ii) the use of the Semantic Web and its corresponding tools and standards.

In both cases, ontology facilitates movement towards unification, thereby further supporting premise

(2a): *ontology provides the features that, in virtue of their absence, make a literature disunified.*

Let us now turn to the final line of support for (2a) in the third goal of ontology.

2.2.4 Third Goal of Ontology: Representing the World

The third goal of ontology is to help us accurately represent the world. Not all ontologies (nor all *ontologists*) are realist(s). It is possible to deny that an ontology does or should represent reality, and thus deny that a goal of ontology building is to represent the world, accurately or otherwise. This subsection defends the adoption of this third goal as a principle of best practice for building or utilizing ontologies. This principle underlies BFO and the OBO Foundry ontologies.¹⁴⁷

2.2.4.1 Knowledge Is Not Reality: Realism and Conceptualism

Before turning to the defense of realism, it is instructive to consider the alternative, sometimes called ‘conceptualism’.¹⁴⁸ According to conceptualism, ontologies do not represent reality, but instead represent our knowledge of reality, typically understood in terms of our *concepts* (hence the label ‘conceptualism’).¹⁴⁹ This was the standard view for some time in ontology building, and many still either subscribe to or implicitly assume it in practice.¹⁵⁰ While conceptualism takes the terms in

¹⁴⁷ See Arp et al. (2015) and The OBO Foundry (2020b). Again, the two lists of the principles are in Appendices B and C.

¹⁴⁸ The discussion here draws on Arp et al. (2015, pp. 5–11), as well as the materials provided in Coursera’s MOOC entitled, “Web of Data” (<https://www.coursera.org/learn/web-data>). Note that ‘conceptualism’ refers specifically to a way of doing ontology and may not correspond to particular philosophical theories of the same name.

¹⁴⁹ Of course, our concepts or knowledge are real entities in the world. This is not a problem for realism and would be covered by a realist ontology of psychology, for instance. The point about conceptualism is that *every* ontology’s terms are intended to represent concepts or knowledge. For instance, in a conceptualist ontology a class like **TUMOR** might be defined as *the term used to refer to an abnormal mass growing inside an organism*. A tumor is not a term.

¹⁵⁰ In fact, it is plausible that the nature of the origin and beginning of both the WWW and the Semantic Web was such that conceptualism was simply the natural view to adopt.

ontologies to represent our knowledge (understood as concepts in people's heads), realism takes them to represent types of entities and relations in reality. So, the relevant question is: Why be a realist?

2.2.4.2 Let's Get Real: Realism Is a Better Standard Practically

Here I provide some positive, practical reasons in favor of adopting a realist principle of ontology building. Before turning to those, consider first that there is some empirical evidence that realism is overwhelmingly the most common position in philosophy concerning the reality of the external world. In a survey of more than 900 philosophy faculty from various universities, David Bourget and David Chalmers found that 81.6% subscribed to some form of non-skeptical realism, while only 9.1% subscribed to some form of either skepticism (4.8%) or idealism (4.3%).¹⁵¹ While this is obviously not conclusive evidence, such a disproportionate distribution of views on this question is at least some *prima facie* evidence in favor of realism.

I refer to these reasons as 'practical' because they promote pragmatic features (interoperability, consistency, scalability) that facilitate unification and ultimately the achievement of the first two goals of ontology, carving up the world (*a la* scientific investigation) and organizing our findings. That is to say, adopting a realist methodology is the most useful way to get the things we want and need out of doing ontology while avoiding common errors that thwart this goal. It is also most congruent with the scientific method as implemented by actual scientists doing science.¹⁵²

First, scientists are realists, at least in practice, because science is an investigation of *reality* and not our *conceptualizations* of reality. Here is Barry Smith and colleagues on this point:

Suppose, again, that Sally attempts to create a representational artifact that makes reference to Tower Bridge by drawing a picture. Our view is that it is here not the mental representation in her head, or the memories in her head, that Sally is trying to draw; rather, it is Tower Bridge itself...

All of this holds true, too, of the representations created by scientists. When constructing such a representation — whether it be a scientific theory presented in a textbook, or the content of a journal

¹⁵¹ See Bourget & Chalmers (2014).

¹⁵² My discussion of this issue of adopting *ontological realism* draws on the work of Barry Smith and colleagues (Arp et al., 2015, pp. 5–11; Ceusters & Smith, 2010; B. Smith & Ceusters, 2010).

article or of a database — the goal is not to represent in a publicly accessible way the mental representations or concepts that exist in the scientists’ minds. Rather, it is to represent the things in reality that these representations are representations of.¹⁵³

In short, since scientists are investigating *reality*, this is what their theories, terms, data, papers, and so forth are referring to. Of course, they must *conceptualize* reality to do so, since this is part of how we access it. However, this does not imply that it is the concepts themselves that are the objects of study.¹⁵⁴ When a biologist writes or says ‘cell’ or ‘metabolism’, they intend to refer to some *actual* biological phenomena out there in the world. Given the purpose of ontology and its connection with scientific investigation, our ontologies should be doing the same.

Second, a realist approach avoids making simple mistakes in our ontologies. While humans may be able to resolve or work around some errors by, for instance, implicitly supplying context or making minor mental adjustments to avoid inconsistencies, computers cannot do this. For lack of a better word, computers are simple-minded. They do exactly what they are told with exactly the instructions and information they are given, as it was given. Importantly, given the immense amount of data out there, we *must* solicit the help of computers if we want to achieve organization, interoperability, shareability, and the like in the most efficient and complete sense possible – that is, unification. So, what are some of these errors that realism avoids?

One obvious mistake that conceptualism (or just a lack of concern for or awareness of a realist approach) leads us into is defining entity types in terms of concepts that are, in fact, not concepts. For instance, as Smith and colleagues point out, the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) defines ‘disorder’ as “a concept in which there is an explicit or implicit pathological process causing a state of disease which tends to exist for a significant length of time

¹⁵³ Arp et al. (2015, p. 10).

¹⁵⁴ This is, of course, leaving aside a discipline like psychology. Again, a realist approach to ontology can accommodate this since it *would* then be concepts – *qua* portion of reality – that are being studied and referred to, at least sometimes. Smith and colleagues make the same point.

under ordinary circumstances.”¹⁵⁵ A quick reflection on this definition, however, will reveal that it is simply confusing *the actual disorder in the world* with *the idea of the disorder in our head*. Put simply, disorders are not concepts; concepts of disorders are concepts.

This mistake is closely related to another that conceptualism tends to lead to, which is called the ‘use-mention error’.¹⁵⁶ This is a failure to recognize or adhere to the difference between *using* a term or phrase, on the one hand, and *mentioning* that same term or phrase (usually with single quotes), on the other. The latter is done in order to refer to the term or phrase itself. Consider the below examples of both use and mention of the term ‘car’:

Use of ‘car’: My car is white and starts by pushing a button.

Mention of ‘car’: ‘Car’ has three letters and starts with a ‘c’.

An example of a use-mention error would be for someone to say, “My car is a three-letter word.” I do not drive a three-letter word to the store. I drive my Subaru, which can be referred to by using the three-letter word ‘car’. The concept-based approach to ontology sets itself up to make such simple (but fundamental) mistakes. By taking the terms in an ontology to refer to concepts or ideas in our heads, the distinction between the entities in the world (such as cars) and that which we use to represent them (such as terms like ‘car’ or our ideas of cars) is blurred or, worse still, altogether lost.

Another mistake conceptualism leads to is making it more difficult to modify or merge ontologies. This is because different individuals, groups, and organizations often conceptualize things differently (one of the factors discussed previously that contributes to disunification). On conceptualism, the focus is on our concepts, and so the fact that conceptualizations differ does not get registered as a cause of concern.¹⁵⁷ However, it is often the case that people are simply thinking of

¹⁵⁵ Arp et al. (2015, p. 9).

¹⁵⁶ See Arp et al. (2015, pp. xix–xx, 9, 11–12).

¹⁵⁷ Strictly speaking, conceptualism should see this as reason to introduce *more* terms and definitions. – as many as there are different conceptualizations.

or talking about the same thing in different ways, and hence, there is often no need for introducing further kinds of entities into the ontology. Moreover, consider trying to merge distinct ontologies, or perhaps data sets tagged with distinct ontologies. Given the multiplication of definitions for the same term, of terms for the same definition, and of terms and definitions for the same entity in reality, this is likely to become either exceedingly laborious and difficult or practically impossible. A realist approach can avoid these problems since the focus is on reality itself rather than the terms we use or how we in fact conceptualize the world. This speaks in favor of realism over conceptualism.

Conceptualism also makes it more difficult to modify an ontology. This is because it focuses on our own ideas or conceptualizations of entities and their relations, which inevitably restricts itself to our current understanding of such entities and relations. Worse still, this makes it so that our ontologies are hardly ever wrong. Suppose an ontology defined ‘water’ as *the concept corresponding to the substance composed of H₂O that fills the rivers, lakes, and oceans*. Since the ontology is only capturing how we conceptualize things, it is hard to see how it could be wrong – *even if we were wrong about the world*. Imagine we discovered that a water-like substance, composed of some other chemical compound, XYZ, actually filled the rivers, lakes, and oceans.¹⁵⁸ Our previous definition could not have been wrong, since it is about how we conceptualize water – note the implication that the pre-H₂O ontology component ‘water’ (had there been one) would not have been wrong either.

Moreover, it is unlikely that our *conceptualization* would suddenly change. Should the conceptualist simply substitute ‘XZY’ for ‘H₂O’ in the definition? This is not, in fact, our conceptualization, and so this solution is anti-conceptualist and starts to suggest (correctly, I think) that the real issue is about capturing *the actual substance* we currently refer to with the term ‘water’. Should the conceptualist introduce another term (such as ‘water_{xyz}’) and a corresponding conceptualism-friendly definition? This could work, if everyone started thinking of the substance that

¹⁵⁸ This thought experiment is inspired by Hilary Putnam’s Twin Earth thought experiment (Putnam, 1975).

fills the rivers, lakes, and oceans as water_{xyz} instead of water. However, it is unclear how long this would take or how ubiquitous the change would be.

The easier and more intuitive solution is to just change the definition of ‘water’ so that it accurately captures the substance it refers to. It is very unlikely that use of the original term ‘water’ would easily shift to some new term ‘ water_{xyz} ’ since language usually changes very slowly. It would be more natural to conclude that we were wrong about water, modify the definition accordingly, and continue using the long-standing term ‘water’ to refer to that same substance we were referring to all along; only now, we would have a more accurate understanding of its nature. People may still reflexively associate H_2O with ‘water’, but they would be wrong (in the sense of accurately representing the world) to do so. This is simply a consequence of doing science. Consider the pre- and post-heliocentric conceptions of ‘sun’ and ‘earth’ or the concept of ‘phlogiston’ in the history of science.

The upshot is that in order for this solution to be appropriate, and I think that it is, we need to assume a realist approach to ontology rather than a conceptualist one. The same goes for avoiding simple errors like confusing entities in reality for concepts in our heads.¹⁵⁹ It also goes for avoiding use-mention errors. When our ontologies are intended to represent reality, we can more easily avoid these mistakes, as well as more easily make adjustments as we learn more about reality and merge ontologies as a means of facilitating interdisciplinary research efforts. When our ontologies are intended to represent our concepts, these mistakes are both easily made and easily missed.

2.2.4.4 Additional Principles of Best Practice

In addition to realism, Smith and colleagues and the OBO Foundry describe and defend a number of additional principles of best practice for building ontologies. The OBO Foundry lists thirteen principles against which any ontology submitted to the Foundry is evaluated, and which the

¹⁵⁹ In conjunction with basic, intuitive axioms like those regarding the *is_a* relation, this also leads to further problematic logical entailments.

Foundry further recommends as good practice for any ontology building, whether it is intended to be a part of the Foundry or not. Smith and colleagues identify eight principles for building ontologies with BFO for specific domains, which have some overlap with the OBO Foundry principles. Furthermore, they identify an additional twenty-five principles for ontology design, which are more specific and related to terminology, definitions, and classification, and which have some overlap both with the OBO Foundry and with much of our discussion of the basics of ontology above.

The basic idea is that, like realism, these principles are taken on board because they both actively promote unification, as well as help to avoid mistakes like those just canvassed that lead to inconsistencies and hinder unification.¹⁶⁰

2.2.5 What the Goals of Ontology Have Taught Us

We have now seen the full support for premise (2a). Ontology provides the features that, in virtue of their absence, make a literature disunified. Given the difficulty of communicating effectively and clearly, the need for ontology is quite ubiquitous, its utility not being restricted to scientific or academic contexts. Simply attempting to understand the world already introduces a core piece of the problem context that calls for ontology. Moreover, since most (if not all) scientific and academic disciplines and their respective research endeavors are faced with large, heterogenous datasets, these disciplines and research endeavors end up on a fast-track to disunification. However, it should now be apparent how ontology can help us to carve up the world, organize our findings, and represent reality in ways that are clearer, more consistent, and better suited towards interdisciplinary research and dialogue. In other words, we have seen how the components and methods of ontology facilitate intra- and interdisciplinary methodological unification.

In the next two sections, I defend premise (2b).

¹⁶⁰ For a fuller explanation and defense of the principles of best practice, see Arp et al. (2015). Their Chapter 1 discusses realism, and their Chapters 3-4 discuss the remaining principles. See also The OBO Foundry (2020b).

2.3 Unification through Ontology: Two Success Stories

Premise (2b) states that ontology has solved problems of disunification for other domains of research which possess no relevant differences from the domain of addiction research. There are two parts to this premise. The first refers to ontology's past successes in combating disunification, and it is fleshed out and defended in the present section. The second utilizes an appeal to analogy, claiming that the field of addiction research is analogous in all the relevant ways to those fields of research helped by ontology, and it is fleshed out and defended in Section 2.4.

I focus here on two examples of ontology's success. The first is the Gene Ontology (GO), widely considered to be one of the clearest success stories for ontology. The second is the Ontology for General Medical Science (OGMS), which is another successful ontology that possesses the additional feature of being specifically tied to medicine. Hence, given its connection to medicine, the OGMS example is more closely connected to the field of addiction research in particular.¹⁶¹

2.3.1 Just GO with It: Ontology for Biology

GO is a reference ontology for the domain of biology, meaning it is “an ontology that is meant to be a canonical, comprehensive representation of the entities in a given domain [biology] that is developed to encapsulate established knowledge of the sort that one would find in a scientific textbook.”¹⁶² In the 1990s, biologists working on completing the genomes of organisms like fruit flies and yeast recognized that surveying the genomic data, comparing data across taxa (such as to the then incomplete human genome), and ultimately storing, describing, and exchanging such data within and

¹⁶¹ Being success *stories* (in the sense that they have narratives, and not in the sense that they are fictions), these examples have histories, background contexts, and a plethora of other details that we do not have the space to get into. Thus, I keep the stories relatively short here. For more on GO, see Dessimoz & Škunca's (2017) anthology, and in particular Chapter 21, “The Vision and Challenges of the Gene Ontology” (S. E. Lewis, 2017). See also Ashburner et al. (2000) and The Gene Ontology Consortium (2019). For more on OGMS, see Arp et al. (2015, pp. 158–167), Ceusters & Smith (2010), Scheuermann et al. (2009), Smith & Ceusters (2015), and the BioPortal page for OGMS: <https://bioportal.bioontology.org/ontologies/OGMS>.

¹⁶² Arp et al. (2015, p. 40).

across disciplinary and institutional boundaries posed a serious challenge.¹⁶³ They faced the same problems described by He and colleagues' five V's of Big Data Analytics (Section 2.2.3.1 above) – in particular, the *volume*, *velocity*, and *variety* of the data being produced made efficient data management, integration, and analysis very difficult.

Suzanne Lewis, a biologist involved in the project, described the problem they faced, nicely capturing the main issues and offering an example of the benefits of finding a solution:

Biologists needed a way of making some sense of the information we were so diligently collecting about genes, both to locate information and to traverse across taxa.

Specifically one slightly obsessive biologist, Michael Ashburner, wanted to classify all fly genes and have the corresponding worm, mouse, human, [and] yeast groups use the same classification scheme...That way, if he found a fly gene involved in a particular process, he could then ask what genes in other taxa are (thought) to be involved in the “same” process, and what insights can be gleaned from its counterpart? We needed a way to describe the attributes of gene products in a rigorous way that would enable biologists to roam the universe of genomes and biology, to explore...[the various] attributes of genes that are of great interest to all biologists. And in an ideal world all biological databases would agree on how such information can be made discoverable and comparable.¹⁶⁴

Given our discussion up to this point, we can see that the context of their investigation of (a biological portion of) reality called for ontology. Thus, GO became the answer – a common, controlled vocabulary for regulating the ways in which researchers across disciplines and institutions, studying various organisms (and their various parts, processes, functions, and so on), described their findings. More specifically, they sought interoperability and unification of their research and disciplines.

2.3.1.1 The Glue of the GO: Some Unifying Features

GO utilized many of the basic components of ontology building discussed above. For instance, unique identifiers (URIs) were used to represent types and classes rather than having labels alone do this, which enabled stable, unambiguous referencing.¹⁶⁵ Recall that this shifts the semantic burden from the label of a class to its definition and associated relations, facilitating easier changes if

¹⁶³ See Lewis (2017) and Bada et al. (2004).

¹⁶⁴ Lewis (2017, p. 292).

¹⁶⁵ See Lewis (2017).

needed and allowing the utilization of multiple labels or names if others are preferred by different researchers or institutions. Again, what ultimately matters are the types and relations being represented, not what we call them. In this vein, GO employed clear, human-readable definitions, as well as synonyms where applicable. As Lewis put it, the goal was “to accommodate every individual researcher by speaking in their particular idiom.”¹⁶⁶ Developers of GO also incorporated various relationships between classes (it had a *graph structure*), including the backbone *is_a* relation, the *part_of* relation, and many others such as *is_regulated_by*, *occurs_in*, *has_scope*, and *starts_during*.¹⁶⁷

These simple components of ontology building already facilitate unification, as has been argued. Consider some examples to see how this is so in the context of GO.

- (i) **Polysemy:** The term ‘alcohol dehydrogenase’ can refer to entirely different types of entities, such as a **PROTEIN** (which *is_a* **INDEPENDENT CONTINUANT** in BFO) or a **PROTEIN FUNCTION** (which *is_a* **DISPOSITION**, and thus a **SPECIFICALLY DEPENDENT CONTINUANT** in BFO).¹⁶⁸ This obviously has the potential for miscommunication and confusion, as well as logical inconsistencies. However, URIs control which entity types (and corresponding definitions and relations) a term or label the URI is associated with is meant to pick out. Researcher A can utter the same words as Researcher B while each means to pick out a completely different type of entity, so long as the data is tagged with the appropriate URI. It matters not what they each *call it*, but only what they each *mean*. GO accommodated polysemy.
- (ii) **Synonyms:** The terms ‘dopaminergic process’, ‘dopamine binding’, ‘catecholamine binding’, ‘dopamine uptake’, and ‘reward process’ might all be used to refer to the same phenomenon of *John’s brain processing a reward*.¹⁶⁹ They may also be intended to provide extra information, such as the specific type of neurotransmitter involved or whether it moves into or out of a cell. Again, the potential for exacerbating confusion and general disunification seems clear. Moreover, as with polysemy, the solution is partially grounded in shifting the semantic burden from linguistic labels to URIs. Moreover, synonyms can be linked to URIs and, where possible, further distinctions and relations between related terms can be specified for added clarity.¹⁷⁰ We might also link ontologies (and so data tagged by them) by indicating that **DOPAMINE** (from the ChEBI ontology)

¹⁶⁶ Lewis (2017, p. 294).

¹⁶⁷ See Ashburner et al. (2000) and The Gene Ontology Consortium (2019).

¹⁶⁸ See Lewis (2017, p. 294).

¹⁶⁹ See Courtot (2016).

¹⁷⁰ For instance, we could specify that **DOPAMINE UPTAKE** *is_a* **TRANSPORT** while **DOPAMINE BINDING** *is_a* **CATECHOLAMINE BINDING** (since **DOPAMINE** *is_a* **CATECHOLAMINE** according to the Chemical Entities of Biological Interest (ChEBI) Ontology). The former (**TRANSPORT**) refers to movement into, out of, or within a cell, while the latter (binding) refers to selective interaction between a molecule and a site on another molecule. See the ChEBI Ontology and its classes here: <https://bioportal.bioontology.org/ontologies/CHEBI>. See GO and its classes here: <https://bioportal.bioontology.org/ontologies/GO>.

participates_in DOPAMINE BINDING and DOPAMINE UPTAKE (both from GO). Further, we can indicate that ‘reward process’ is a synonym that might be used to refer to one or all of these more specific processes.

- (iii) **Who’s in Charge?** Here are some facts: some terms are polysemous; some researchers or groups have local (lab- or experiment-specific) goals; some researchers or groups have idiosyncratic vocabularies or terms; some statements about biology are contentious while others are not or are less so. This has the potential to facilitate disunification since researchers might talk past one another, create unshareable or unscalable databases, or otherwise utilize methods or participate in research that resists interoperability. GO has resisted this would-be disunification through a number of pro-collaboration features, such as: involvement of the biological community; having clear, explicit goals; limiting its scope; possessing a simple, intuitive structure (a directed, acyclical graph); and undergoing continuous evolution and active curation.¹⁷¹ In short, GO operates under the explicit intention that interoperability of biological research and data is the goal, and it incorporates the very researchers gathering the data to achieve this goal rather than simply dictating what is to be from the outside.

GO began its early years with a few thousand terms and currently has more than 44,000 validated terms representing entity types within or related to its three sub-domains: *biological processes*, *molecular functions*, and *cellular components*.¹⁷² More than 1.5 million gene products and over 4,600 species have been annotated using GO, and further GO boasts upwards of 8 million data annotations in total. What is more, GO has never been (and was never *intended* to be) a static artifact. The number of scientific publications with GO annotations has increased by 10,000 (a nearly 7% jump) in just the last two years. While the total number of terms has remained somewhat steady over those two years, those maintaining GO continually work to modify the vocabulary as needed per feedback from the relevant scientific communities and subject matter experts. They render terms obsolete, create them anew, or merge them between successive versions of GO as appropriate. Recall that this continued development is occurring some *25 years* after its original inception.

¹⁷¹ See Bada et al. (2004).

¹⁷² These and the following statistics are from GO’s homepage: <http://geneontology.org/stats.html>. These sub-domains also serve as the three root nodes of GO, each connecting up to its corresponding parent class in BFO: **PROCESS**, **FUNCTION**, and **MATERIAL OBJECT**, respectively.

2.3.1.2 GO-ing Towards Unification

As Smith and colleagues note, the problem GO was originally designed to address is “common across the whole of science: where multiple disciplinary groups are involved in the study of some scientific phenomenon of interest each will likely have its own idiosyncratic vocabulary.”¹⁷³ The solution GO offered was to create a formally structured, human-readable vocabulary, based on input from the researchers themselves, which was *logically consistent*, *shareable*, and *controlled*. This vocabulary was then systematically used “by literature curators to describe experimental data appearing in published papers...[making the data] more easily retrievable and combinable, in ways that overcome the problems caused by multiple conflicting vocabularies.”¹⁷⁴ It is no wonder that GO is widely considered the most successful ontology to date.

The success of GO is our first line of support for the first part of premise (2b), that ontology has solved problems of disunification for other domains of research. Let us now turn to the second.

2.3.2 Two Thumbs for OGMS: Ontology for General Medicine

Like GO, OGMS is also a reference ontology. Smith and colleagues describe it as being “designed to provide a formal, explicit, nonredundant, and unambiguous representation of clinical terms” so as to address and avoid the inconsistencies, vagueness, and discipline-dependence of clinical terminology.¹⁷⁵ Rather than being an ontology of disease, OGMS “provides the terminological core of a general theory of disease and formal definitions for terms widely used in clinical encounters to describe different aspects of disease.”¹⁷⁶ Here is its description from the OGMS homepage:

The Ontology for General Medical Science (OGMS) is an ontology of entities involved in a clinical encounter. OGMS includes very general terms that are used across medical disciplines, including: ‘disease’, ‘disorder’, ‘disease course’, ‘diagnosis’, ‘patient’, and ‘healthcare provider’. OGMS uses the

¹⁷³ Arp et al. (2015, p. 60).

¹⁷⁴ Arp et al. (2015, p. 60).

¹⁷⁵ Arp et al. (2015, p. 159). For the seminal papers that served as the foundation for the initial development of OGMS, see Smith et al. (2007) and Scheuermann et al. (2009).

¹⁷⁶ Arp et al. (2015, p. 159).

Basic Formal Ontology (BFO) as an upper-level ontology. The scope of OGMS is restricted to humans, but many terms can be applied to a variety of organisms. OGMS provides a formal theory of disease that can be further elaborated by specific disease ontologies.¹⁷⁷

Thus, from this foundational framework, additional application ontologies can be and have been developed as extensions of OGMS to cover a range of different diseases, disease families, or medically relevant subjects. Some examples are the Drug Ontology (DrOn), the Infectious Disease Ontology (IDO), the Mental Disease Ontology (MFOMD), the Ocular Disease Ontology (ODO), the Sleep Domain Ontology (SDO), and the Vital Sign Ontology (VSO), among several others.¹⁷⁸

2.3.2.1 Some Unifying Features of OGMS

OGMS helped to remedy disunification in the context of medical research. First, OGMS shares the basic unification-promoting features of an ontology that we saw with GO that serve to embody the principles of best practice underlying BFO. These include the use of URIs, a backbone *is_a* hierarchy, Aristotelian definitions, realism, a commitment to collaboration, and so on. Second, OGMS has been extended with further ontologies that cover more domain- and topic-specific areas of study in the medical field. In this sense, it serves as a kind of top-level ontology for the medical domain, and hence, ensures a common, controlled vocabulary is utilized across the extension ontologies. The continued success and reuse of OGMS is therefore unification-promoting in itself.

Let us look at a few examples of unification in the context of OGMS.

- (i) **More Polysemy and Synonyms.** As with biological terms, the same medical or clinical term can have multiple different uses, while different ones can be used to mean the same thing. For instance, ‘diabetes’ might be used to refer to a diabetic’s condition, as when we say, “Graham’s diabetes was discovered late.” Alternatively, it can be used to refer to the entire course of acquiring, living with, and treating diabetes and its symptoms, as when a doctor says, “Ruth, diabetes is not an easy thing to endure.” On the other hand, one might use ‘diabetes’, ‘diabetes mellitus’, ‘Type 2 diabetes’, or ‘adult-onset diabetes’ to refer to one and the same condition. Like GO, OGMS utilizes URIs to designate particular entity types (and corresponding definitions), as well as synonyms to designate that a class’s label can be substituted for another without loss of meaning. For the case

¹⁷⁷ See the OGMS homepage at: <https://github.com/OGMS/ogms>.

¹⁷⁸ See Arp et al. (2015, p. 162), and the OGMS homepage: <https://github.com/OGMS/ogms>.

of polysemy, consider the distinction in OGMS (displayed in Figure 7 below) between **DISEASE** (OGMS:0000031, which *is_a* **DISPOSITION**, and thus a **CONTINUANT** in BFO) and **DISEASE COURSE** (OGMS:0000063, which *is_a* **PROCESS**, and thus an **OCCURRENT** in BFO). This disambiguates the two distinct uses of ‘diabetes’ above.

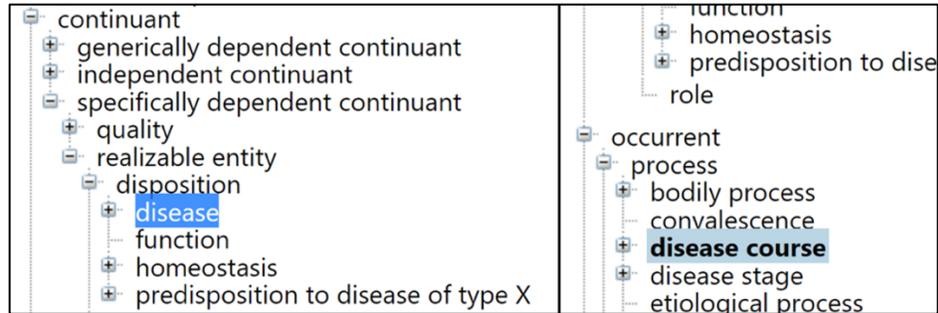


Figure 7: **DISEASE** and **DISEASE COURSE** Are Distinct Classes in OGMS

- (ii) **Disease, Disorder, and Dysfunction.** A related but distinct issue in the medical context is the fact that ‘disease’, ‘disorder’, and ‘dysfunction’ are sometimes used interchangeably. For instance, the *International Classification of Diseases, 11th Edition* (ICD-11) does not explicitly define these terms but classifies **MALE ERECTILE DYSFUNCTION** as a *subtype_of* **MENTAL, BEHAVIORAL, OR NEURODEVELOPMENTAL DISORDERS** (via the intermediary class **SEXUAL DYSFUNCTION**). It also classifies **SAPHO SYNDROME** as *subtype_of* **AUTOINFLAMMATORY DISORDERS**, which *is_a* **DISEASE OF THE IMMUNE SYSTEM**. This is one of the most central, long-standing classification systems in the medical field classifying dysfunctions as disorders, syndromes as disorders, and disorders as diseases. Hence, it is unsurprising that there is varied use of and terminological confusion over these terms. OGMS, on the other hand, clearly distinguishes **DISEASE** (a BFO **DISPOSITION**) from **DISORDER** (a BFO **MATERIAL OBJECT** that is a *clinically abnormal part of an ORGANISM*). Figure 8 below illustrates how OGMS distinguishes these and related entity types (disambiguating the terms), and the relations they stand in.

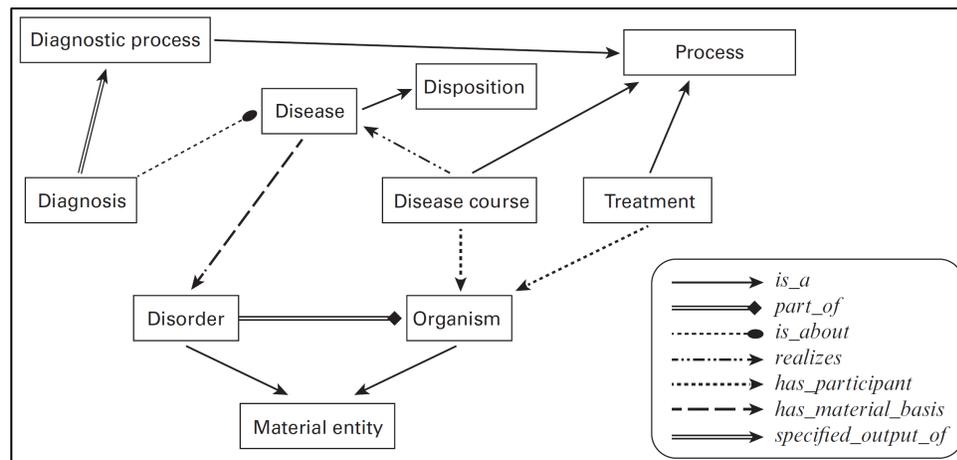


Figure 8: Relations between **DISEASE**, **DISORDER**, and Related Entity Types¹⁷⁹

¹⁷⁹ This diagram is taken from Arp et al. (2015, p. 164).

- (iii) **Disease vs. Evidence for Disease.** Figure 8 above also shows that OGMS helpfully distinguishes important entities and processes on the side of the patient (such as **DISEASE** and **DISORDER**) from those on the side of the clinician (such as **DIAGNOSIS**). A diagnosis of a disease is not the disease itself. It is an **ASSERTION** which is intended to be *about* a disease.¹⁸⁰ OGMS also distinguishes between **SIGN** and **SYMPTOM**, understood as involving epistemic components like *being judged clinically significant by the clinician* or *being hypothesized to be a manifestation of a disease*, respectively.¹⁸¹ These are basic distinctions and so this may seem obvious. Nonetheless, missing or disregarding them does occur (as the ICD-11 example above demonstrated), and doing so can result in problems such as confusing *evidence for a disease* with *the disease itself* (as some addiction researchers do). Figure 9 below helps to illustrate in even more detail some of these central distinctions that combat confusion, inconsistencies, and overall disunification.

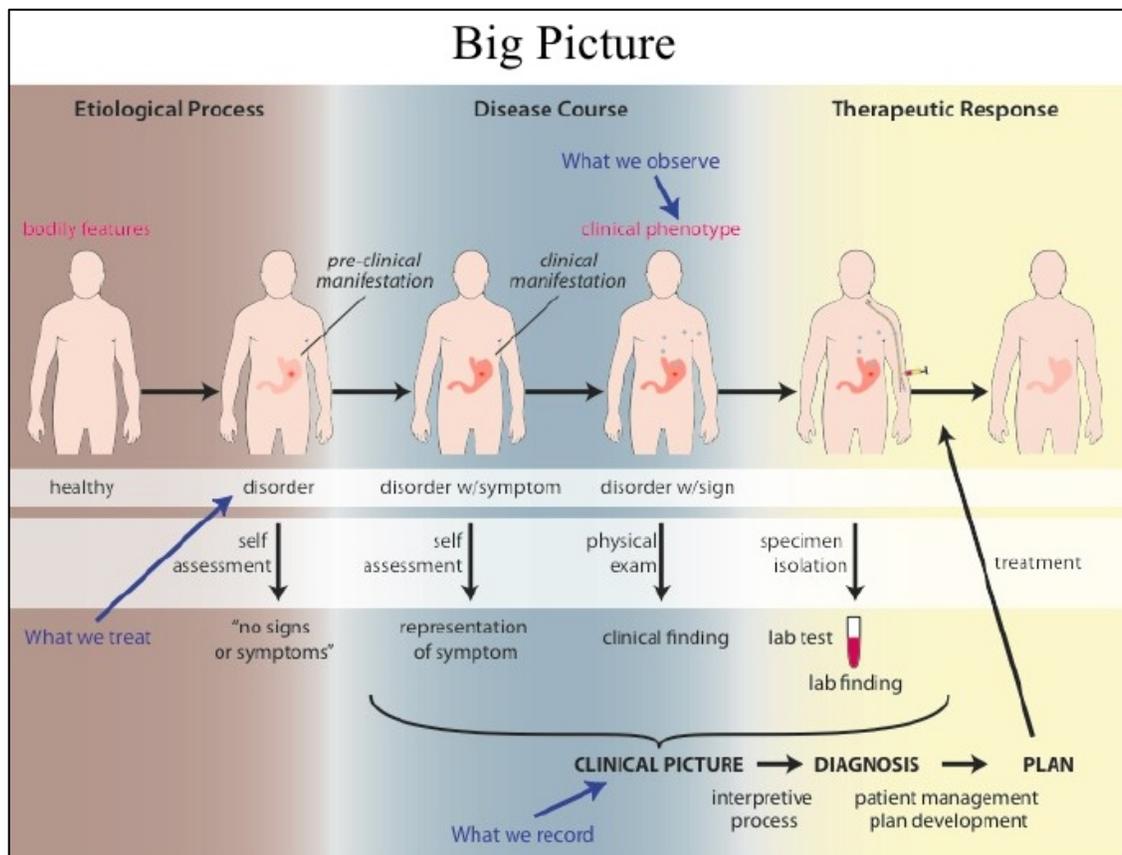


Figure 9: Components of a Disease Course¹⁸²

¹⁸⁰ See Limbaugh et al. (2020) for a careful discussion of these entities in the context of developing and applying the Cognitive Process Ontology (CPO) to cover the concept of ‘warranted diagnosis’.

¹⁸¹ See OGMS classes and definitions here: <https://bioportal.bioontology.org/ontologies/OGMS>.

¹⁸² This diagram is taken from <https://www.slideshare.net/BarrySmith3/basic-formal-ontology-a-common-standard>.

2.3.2.2 OGMS and Unification

As with GO, OGMS was created to rectify intra- and interdisciplinary disunification, such as lacking a shared vocabulary and the inability to effectively communicate or efficiently manage data. Its target was general medical science rather than biology, but the problems and solutions were the same. Barry Smith and colleagues describe this in a seminal paper initiating the development of OGMS:

Effective knowledge representation requires the use of standardized vocabularies to ensure both shared understanding between people and interoperability between information systems. Unfortunately, many existing biomedical vocabulary standards rest on incomplete, inconsistent or confused accounts of basic terms pertaining to diseases, diagnoses, and clinical phenotypes... The effective study of [the ways disease correlates with genetic and environmental variables] requires clinical research to be applied to ever larger pools of subjects drawn from geographically separated populations in multi-institution studies, requiring that the healthcare institutions involved embrace common standardized terminologies in capturing and sharing their data. The definitions presented here are designed to provide the resources in terminology and disease classification to support such standardization.¹⁸³

OGMS has many successful extensions, and so it is certainly supporting such standardization. Hence, OGMS is our second line of support for the first part of premise (2b), that ontology has solved problems of disunification for other domains of research. Let us now turn to the defense of the second part of (2b), that with respect to solving problems of disunification, biology and medicine possess no relevant differences from the domain of addiction research.

2.4 Applying Ontology to Addiction Research

The remaining defense of premise (2b) is really quite simple. First, it will be useful to recall from Chapter 1 my reconstruction of Janna Hastings and colleagues' argument for the expected success of using ontology in addiction research. Their argument was formulated roughly as follows:

- (P1) Implementing realist, BFO-conformant ontologies into domains like biology and medicine helped solve their problems of disunification;
- (P2) With respect to solving their problems of disunification, there is no relevant difference between these domains and the domain of addiction research;

¹⁸³ Scheuermann et al. (2009, p. 116).

- (P3) If (P1) and (P2), then implementing realist, BFO-conformant ontologies into addiction research will help solve its problems of disunification;
- (C1) Hence, implementing realist, BFO-conformant ontologies into addiction research will help solve its problems of disunification. [(P1)-(P3)]

The argument is similar to the one I have been defending here. Indeed, the preceding sections on GO and OGMS support (P1) of the above argument, making good on my Chapter 1 promise to expound on the support for this premise. The present section concerns the extension of ontology into addiction research, and so a defense of (P2) is of interest. This is the same as the *no difference* component of my premise (2b), and I turn now to defending it.

2.4.1 What's in a Discipline? The No Difference Argument

With the fundamentals of ontology under our belts, it is much easier to see why (P2) is hardly contentious at all, despite being the most contentious premise in the above argument. Investigating the world (doing science, broadly speaking), effectively organizing our findings, and inevitably having to compile and manage large heterogeneous datasets all call for ontology. This is especially so when the data fits He and colleagues' five V's discussed above – it is produced at high *volume*, with high *velocity*, from a *variety* of sources, and needs to have certain degrees of *veracity* and *value* to be useful. Addiction research certainly meets these criteria.

A quick search of a few addiction terms on the academic search engines PsychNet (PN) and ScienceDirect (SD) produced the following quantities of results: 121,260 (PN) and 124,833 (SD) for 'addiction'; 26,926 (PN) and 337,495 (SD) for 'substance use disorder'; and 31,737 (PN) and 211,378 (SD) for 'alcohol dependence'. The results are from a wide variety of journals as well, demonstrating the interdisciplinary nature of addiction research, and thus the need for interoperability. These included *Behavioral Brain Research*, *Social Science and Humanities Open*, *Archives of Psychiatric Nursing*, *Diabetes & Metabolic Syndrome*, *Pediatric Clinics of North America*, *International Journal of Drug Policy*, *Computers in Human Behavior*, and *Technology in Society*, to name just a handful. Furthermore, addiction is incredibly

consequential, and so the data need to be veracious and valuable. Consider that roughly half (46%) of Americans report having a family member or close friend with an addiction.¹⁸⁴ Additionally, in the U.S. addiction exacts over \$740 billion annually¹⁸⁵ and deaths from drug overdose have been rising annually since 2002, exceeding 50,000 in 2015.¹⁸⁶

It is clear that addiction research needs ontology. Again, the issue is providing a common, controlled vocabulary for talking about the relevant kinds of entities and relations, and a shared methodology for organizing and using that vocabulary (and the data it is used to tag). Because of this, it is hard to see how *any* domain of research would be relevantly different from those where GO and OGMS were successful. More importantly (and more clearly), *addiction research in particular* shares all the qualities that made biology and medicine perfect candidates for the introduction of ontology as a means towards unification. For instance, much addiction research falls directly within these domains. Hence, premise (2b) is well-supported: ontology has solved problems of disunification for other domains of research which possess no relevant differences from the domain of addiction research. Here again is the full defense of premise (2) of my overarching argument:

- (2a) Ontology provides the features that, in virtue of their absence, make a literature disunified; [Section 2.2]
- (2b) Ontology has solved problems of disunification for other domains of research which possess no relevant differences from the domain of addiction research; [Sections 2.3-2.4]
- (2c) If (2a) and (2b), then ontology can help to solve the addiction literature's problem of disunification; [with (1) from the overarching argument, trivially true]
- (2) Hence, ontology can help to solve the addiction literature's problem of disunification. [(2a)-(2c)]

¹⁸⁴ See Gramlich (2017). This also says nothing of those merely acquainted with an addict (coworkers, classmates, neighbors, friends of friends). Controlling for sex, political affiliation, or race revealed no statistically significant differences. In other words, the prevalence of addiction is immune to who one is, where they are from, and so on.

¹⁸⁵ See the National Institute on Drug Abuse (NIDA) summary of a 2007-2010 Department of Health and Human Services report: <https://www.drugabuse.gov/related-topics/trends-statistics>.

¹⁸⁶ See the NIDA summary of the Center for Disease Control and Prevention database on drug-related death rates: <https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates>.

2.4.2 Prototype Addiction Ontologies

Chapter 1 discussed the recent call for the implementation of ontology to addiction research. More than this, Janna Hastings and colleagues have twice attempted to develop such an ontology. The first attempt came some eight years ago and mostly focused on addiction itself (*qua* entity). It used the Mental Disease Ontology (MFOMD), an extension of the Mental Functioning Ontology (MF), as the foundation.¹⁸⁷ This is because they understood **ADDICTION** as a *subtype_of* **MENTAL DISEASE**. The second attempt began alongside West, Hastings, and colleagues' recent call for getting ontology into addiction research.¹⁸⁸ This more recent attempt is a wider-reaching and more comprehensive approach than the first, focusing on the many, often disparate components of the larger phenomenon of addiction (such as objects of addiction like cocaine, addiction treatments, and so on). This ontology of addiction would utilize a wide variety of existing ontologies to incorporate its many components:

Addiction involves many constructs that are not addiction-specific, and the series will cover these constructs and associated ontologies to the extent that they are relevant for our field.¹⁸⁹

Hence, addiction research is an entirely appropriate target for the use of ontology as a remedy for disunification, and there have been at least two serious attempts to construct such an ontology for that purpose. Let us now turn to a summary of the lessons from this chapter.

2.5 Lessons from Chapter 2

In this chapter, I have explained and defended the methodology of ontology and the appropriateness of its implementation into addiction research. We began with the context out of which the need for ontology arises, and saw that this context was quite ubiquitous and, given the general difficulties of communication, quite unsurprising. This context boasted two central features: scientific

¹⁸⁷ See Hastings et al. (2012b).

¹⁸⁸ See West, Christmas et al. (2019a) and West, Marsden et al. (2019b).

¹⁸⁹ West, Marsden et al. (2019b, p. 955). See also the homepage for the ontology here: <https://addictovocab.org>.

investigation (broadly understood) and heterogenous data. We also needed a sufficiently robust understanding of the fundamentals of ontology in order to properly understand its later applications. Section 2.2 provided a survey of these fundamentals as they related to three important goals of ontology: carving up the world, organizing our findings, and representing reality.

Concerning carving up the world, ontology helps us to distinguish types of entities and the relations they stand in with one another. We learned of continuants, occurrents, and some of their subtypes that constitute the central domain-neutral ontology, BFO. We learned of various relations, but most importantly the hierarchical *is_a* relation serving as the backbone taxonomic structure of any ontology. Finally, we learned of Aristotelian definitions, universals and defined classes, and the idea that ontologies provide a *shared, structured vocabulary*.

Concerning organizing our findings amidst a sea of computational, heterogenous data, ontology helps both through the very practice of classification that it embodies, as well as through the Semantic Web and its associated tools.

Concerning representing reality, we saw that ontology ought to adhere to realism over its (perhaps more frequently subscribed to) rival, conceptualism. I argued that a realist framework is the better choice practically. Ontological realism is much more closely aligned with actual science, it helps to avoid simple yet fundamental mistakes like the use/mention error, and it facilitates the merging and modifying of data and ontologies.

With our ontology toolkit on the table, the latter half of the chapter took us through two examples of the success of ontology in bringing unification to two fields of research. GO and OGMS combatted the errors, confusion, inconsistencies, and general disunification brought on by things like the polysemous and synonymous use of terms and the idiosyncratic methods and vocabularies of individual researchers, groups, disciplines, and institutions. Finally, the penultimate section made it

clear that there is no difference between these fields unified by ontology and virtually any field of study, let alone the domain of addiction research.

In the next chapter, the focus shifts from methodology to substantive content as I lay out and defend the components of my dispositional account of the nature of addiction. We will also see in Chapter 3 why the content I offer is itself a way to provide a further, distinct kind of unification to the literature. Let us turn now to the substance.

Chapter 3: Towards a Dispositionalist Account of Addiction

3.1 Introduction and Chapter Road Map

This chapter introduces and begins the defense of the substance of my account. Its focus is setting the stage for the fuller defense of premise (3) of my overarching argument, which comes in the next chapter. Premise (3) states that the dispositionalist account of addiction is true and provides still further unification to the literature (over and above that provided by the implementation of ontology). Here is the argument for premise (3):

- (3a) Ontology cannot *fully* unify the literature (that is, there is a further, distinct sense of ‘unification’ that ontology alone cannot provide);
- (3b) The dispositionalist account of addiction best explains key phenomena surrounding addiction;
- (3c) The dispositionalist account of addiction captures other competing theories and can explain why their disagreements are sometimes only apparent;
- (3d) The dispositionalist account works within a realist, BFO-conformant ontology;
- (3e) If (3a)-(3d), then the dispositionalist account of addiction is true and provides still further unification to the literature;
- (3) Hence, the dispositionalist account of addiction is true and provides still further unification to the literature. [(3a)-(3d)]

In Chapter 4, I will defend premises (3b)-(3e) of this sub-argument. Here, the stage-setting will consist of three components: an explanation and initial defense of premise (3a); an introduction to dispositions; and an explanation of the core of the dispositionalist account of addiction defended here. Hence, the present chapter is structured as follows.

In Section 3.2, I explain and defend premise (3a). This will lay the foundation for the dispositionalist account and, thus, for defending premise (3). The defense is *initial* only in the sense that to fully appreciate the reasoning behind (3a), we will need to take the full journey through premises (3b) and (3c) in Chapter 4. However, doing so with at least a general idea of what (3a) is getting at will be important.

In Section 3.3, we will be introduced to the world of dispositions. I will highlight the features of dispositions that will be most relevant to the account defended here. It is crucial to understand dispositions and how they work as they are central to my account of addiction and my overall thesis.

In Section 3.4, I present and expound on the core dispositionalist elements of the account I defend, finalizing the stage-setting. That is, with premise (3a) behind us and the world of dispositions at our disposal, I fill in the content of my own dispositionalist account of addiction. This will determine what we have to work with going forward.

Section 3.5 recaps this chapter's arguments before defending premises (3b)-(3e) in Chapter 4.

3.2 The Limitations of Ontology's Unifying Power

This section explains and defends premise (3a) of the above argument. The preceding chapter carefully took us through the basics of ontology, as well as the argument that the methods of ontology facilitate unification. Thus, one might wonder what a particular account of the nature of addiction like the one defended here is supposed to do for us beyond the unifying power of ontology defended in Chapter 2. This section sets out the beginning of an answer to this question. The rest of the answer comes once we have a fuller defense of the account. This is because part of that fuller defense itself will explain some of what the account can do in terms of unifying the literature in a way that is distinct from the unification provided by ontology.

Here is premise (3a) again: ontology cannot *fully* unify the literature (that is, there is a further, distinct sense of 'unification' that ontology alone cannot provide). The present task is an elucidation of this distinct sense of 'unification' and an explanation of why ontology cannot provide it.

3.2.1 Two Senses of 'Unification': Methodological and Substantive

We should be somewhat familiar with at least one kind of unification by now. I will only remind the reader that it pertains to providing a shared, controlled vocabulary so as to ensure a

common understanding of the meaning of terms that represent types of entities and relations within some domain. In turn, this helps to avoid conceptual confusions, verbal disagreements, and simple mistakes like use-mention errors. It also facilitates more effective and efficient communication and data management. Chapter 2 taught us how ontology helps to secure this form of unification, which is *methodological* in nature.

Rather than focusing on methodology, the second kind of unification is *substantive* in that it pertains to the unification of the substance (or content) of different accounts. Methodological unification allows different (often competing) accounts about some phenomenon to effectively interact independently of the content of the respective views. Substantive unification, on the other hand, is about attempting to bring the content of different accounts together under a single, unified framework. A set of views in a literature are methodologically unified just in case, roughly, there exists some appropriate means for successful interaction and understanding among the proponents of those views – such as a common vocabulary, principles for properly controlling it, and so forth. A set of views is substantively unified just in case, roughly, there is some account the contents of which can be said to capture the others. Let me say a bit more about how I understand ‘capture the others’ to help further elucidate this idea.

In the well-being literature, Michael Bishop has proposed what he calls the ‘network theory of well-being’, and it is an example of an attempt to captures other views on offer – that is, to provide what I am calling ‘substantive unification’.¹⁹⁰ Some well-being researchers, such as positive psychologists, focus on and produce accounts that are aimed solely or primarily at positive experiences and attitudes. Others, such as sociologists or social psychologists, focus on positive relationships; neuroscientists and evolutionary biologists often focus on neurological mechanisms, brain states, and

¹⁹⁰ See Bishop (2015). Let us set aside whether his view is true and whether he actually succeeds in unifying other accounts. Instead, just focus on the idea behind his project.

neurotransmitters; philosophers often focus on desire-satisfaction and objective goods; economists and policymakers often focus on income, education level, divorce rates, and other socioeconomic variables; and other researchers from other disciplinary backgrounds may focus on still other features. Bishop proposes that much of this research is on the right track, but ultimately each is insufficient.

It is not that well-being is none of these things, or any one of them, or even just an aggregate of all of them. Instead, he thinks that to be in a state of well-being is to instantiate a certain arrangement of dispositions and to be situated in what he calls a “positive causal network.”¹⁹¹ Positive experiences, for Bishop, are not *what well-being is*, though someone with well-being will certainly be disposed towards such experiences. Nor is it *merely* about brain states, or *merely* about relationships, or satisfying desires, or a certain income level. Still, well-being will tend to incorporate some or all of these things, to varying degrees, in the positive causal network within which someone with well-being is situated. Networks can vary, perhaps consisting of weaker positive relationships (if a person is, say, more introverted) but stronger dispositions toward positive feelings and satisfied desires. The point is that Bishop’s theory is meant to capture what is right about the competing views, and to explain other relevant aspects that might be related but non-essential to well-being. For instance, a positive causal network *can* lead (perhaps even *usually* leads) to a counterfactually better income or education level, or perhaps make it more likely. But this would be a *typical effect* of well-being, and not necessarily essential to it. In this way, Bishop’s network theory is unifying because the content of that theory is meant to capture what is right and explain what is wrong (but relevant) about the alternative views.

Analogous examples can be found in physics. Consider Newton’s *first great unification* in which his theory of gravity unified theories of the behavior of earthly bodies, on the one hand, and celestial bodies, on the other.¹⁹² Consider also Maxwell’s *second great unification* in which his theory of

¹⁹¹ See Bishop (2015, Ch. 3).

¹⁹² See the Wikipedia page on Newton’s unification here: [https://en.wikipedia.org/wiki/Unification_\(physics\)](https://en.wikipedia.org/wiki/Unification_(physics)).

electromagnetism unified theories of magnetism, electricity, and light, showing that they were all “different manifestations of the same phenomenon.”¹⁹³ Finally, consider *unified field theory* in physics, which is “an attempt to describe all fundamental forces and the relationships between elementary particles in terms of a single theoretical framework.”¹⁹⁴ More specifically, it is the attempt to formulate a single field theory that can capture Maxwell’s field theory of electromagnetism, Einstein’s field theory of gravitation (general relativity), and any other theory attempting to explain fundamental forces or relationships between fundamental particles, such as quantum theory. If unified field theory were to be successful, it would explain the phenomena captured by these distinct field theories as aspects of a single fundamental field.¹⁹⁵

Hopefully the idea is clear. Substantive unification is about bringing distinct and sometimes competing theories into (or under) a single framework. This might be by unifying many theories of X . For instance, “ X is p ,” “ X is q ,” “ X is r ,” and “ X is s ” might be unified by a theory holding that p , q , r , and s are all e or parts of e – hence, “ X is e , and e captures p - s .” Alternatively, it might be by unifying theories about seemingly disparate phenomena surrounding or related to X . For instance, consider theories holding that “ X involves p ,” “ X involves q ,” “ X involves r ,” and “ X involves s ,” but where each holds that p , q , r , and s are distinct and so require different perspectives and approaches which do not allow for a single theory (this should elicit memories of the elephant metaphor). Yet a theory might unify these factors by holding that e can account for p , q , r , and s despite the fact that different perspectives and approaches are appropriate – perhaps e is accessible from each perspective or approach, for instance. Hence, we could say that “ X is e , and e captures p - s .”

¹⁹³ See the Wikipedia page on Maxwell’s theory here: https://en.wikipedia.org/wiki/James_Clerk_Maxwell.

¹⁹⁴ See ‘unified field theory’ in the Encyclopedia Britannica here: <https://www.britannica.com/science/unified-field-theory>.

¹⁹⁵ Apparently, this did not work. See the Britannica entry. But the point about what substantive unification is stands.

My contention is that the dispositional account of addiction will do each of these to some degree. This is because differing accounts in the literature come in both types:

- (i) “addiction is p ” vs. “addiction is q ” vs. ...; and,
- (ii) “addiction involves r ” vs. “addiction involves s ” vs. ... (where r, s , etc. purportedly call for distinct perspectives).

Now, with substantive unification on the table, let us briefly turn to why ontology cannot provide it.

3.2.2 Accepting AAA Would Lock Ontology Out of Substantive Unification

We were introduced to the Semantic Web, the home of ontologies, in Chapter 2. As the reader will recall, this cousin of the World Wide Web made it possible to link and, importantly, represent and describe individual datum or facts, as opposed to entire documents containing data and facts. One important feature of both Webs, though, is the principle sometimes referred to as ‘the triple A principle’ (or ‘AAA’). Here is Dean Allemang and James Hendler on this principle:

The Web is the ultimate example of the warning *caveat emptor* (“Let the buyer beware”). This feature of the Web is so instrumental in its character that we give it a name: the *AAA Slogan*: ‘**A**nyone can say **A**nything about **A**ny topic.’¹⁹⁶

Of course, on the WWW this translates to the (probably familiar) idea that *anyone* can create a web page that says *anything* they want it to say, and about *anything* they want to say it about.

If, for instance, Quinton Anonymous wants to create a webpage explaining that J.P. Morgan and the Rothschilds planned to and then sank the Titanic in order to kill a large, influential group who opposed the Federal Reserve, then QAnon (as we might call this person) can do this.¹⁹⁷ Again, buyer (or web browser) beware. It is our responsibility as readers of web content to sort out the good from the bad, the true from the false, and so on. Likewise, in ontology, this translates to the (probably less familiar) idea that *anyone* can create an ontology that represents *anything* they want it to represent, and in *any way* they want to represent it. If, for instance, Jake wants to create an ontology asserting (or with

¹⁹⁶ Allemang & Hendler (2011, p. 6), emphasis in original.

¹⁹⁷ This is based on a real conspiracy theory. See Bird (2015) and Trickey (2018).

the representation) that **SALAD** *is_a* **SANDWICH** *that has no bread*, but then classifies **CAESAR SALAD** as a *subtype_of* the class **SOUP**, then he can do that. Ontologists beware.

As the authors note, while this principle of free expression was influential in allowing the two Webs to take off and ultimately become successful platforms, it clearly has its drawbacks. Indeed, AAA was directly influential in producing the explosion of ontologies that led the OBO Foundry and Barry Smith and colleagues to adopt, refine, and defend the realist methodology and the principles of best practice discussed in Chapter 2. AAA is somewhat of a double-edged sword in this way. It is foundational to RDF, the highly expressive description framework underlying the Semantic Web, but it can also allow ontologies to become counterproductive to the goal of unification that they were intended to help achieve. Here is Allemang and Hendler again on this point:

[AAA] also means that the Web is like a data wilderness—full of valuable treasure, but overgrown and tangled. Even the valuable data that you can find can take any of a number of forms, adapted to its own part of the wilderness...[The] Web has no gatekeeper. Anything and everything can grow there. A distributed web of data is an organic system, with contributions coming from all sources.¹⁹⁸

Consider further a related principle of best practice, *perspectivalism*, which states that there are multiple accurate descriptions of reality. In other words, reality can be described from different levels of granularity, and an ontology should make no commitments about which is the *true* perspective, or about any level of granularity being *more real* than others. Instead, it should treat each perspective or level of granularity as representing reality, and each perspective's corresponding entity types as genuine in their own right – period.¹⁹⁹ Smith and colleagues provide a nice reminder of this concept:

Perspectivalism flows from the recognition that reality is too complex and variegated to be embraced in its totality within a single scientific theory. It amounts to the principle that two distinct scientific theories may both be equally accurate representations of one and the same reality.²⁰⁰

¹⁹⁸ Allemang & Hendler (2011, p. 7).

¹⁹⁹ Recall that this does not mean that AAA means that anything anyone says about anything is *correct* in the sense of *accurately representing the world*. This is why *fallibilism*, the *open world assumption*, and other such principles are just as important. It simply means that, as a starting point, we should be granularity pluralists, accepting that there really are different levels of reality that correspond to the various perspectives taken by distinct disciplines. Roughly, this makes ontologies work much better. Hence, the weaker version is that, for pragmatic reasons, we should just *treat* levels of granularity this way.

²⁰⁰ Arp et al. (2015, p. 44).

Relatedly, and more specifically on the point of AAA applying to building ontologies, Allemang and Hendler add the following:

The Semantic Web is often mistaken for an effort to make everyone agree on a single ontology— but that just isn't the way the Web works. The Semantic Web isn't about getting everyone to agree, but rather about coping in a world where not everyone will agree, and achieving some degree of interoperability nevertheless. There will always be multiple ontologies, just as there will always be multiple web pages on any given topic. The Web is innovative because it allows all these multiple viewpoints to coexist.²⁰¹

Putting all of this together, it is clear why ontology, in principle, cannot provide substantive unification. Its methods (assuming best practice) rule out taking a stand on any one position or theory about *X*. “Come one, come all,” we might say. To be sure, this is certainly not a *knock* against ontology, nor against the principles of best practice Smith and the rest defend. On the contrary, Chapter 2 showed us that ontologies built according to these principles, *perspectivalism* included, have been enormously successful. The upshot is only that, given the *methodological* nature of ontology, and given principles like AAA and *perspectivalism*, the unification facilitated by implementing ontology could not be anything else but methodological.

3.2.3 If Not Ontology, then What?

From the preceding discussion, it follows that there is a distinct kind of unification, substantive unification, and ontology by itself cannot provide this unification to a domain of research. In other words, it follows that premise (3a) is true. This is what I meant in saying that ontology cannot *fully* unify the literature.

If ontology cannot provide substantive unification to the addiction literature, then what can? The dispositionalist account of addiction can do this, aided in large part by its considerable explanatory power. First, though, we need an understanding of dispositions. I am defending a *dispositionalist* account of addiction, after all, and so some foundational components must be laid down and clarified. What

²⁰¹ Allemang & Hendler (2011, p. 9).

are dispositions? How do they work? Are there different kinds? Answering these questions is the task of Section 3.3.

3.3 The Basics of Being Disposed

To better understand the account defended here as we move forward in the project, the reader must be armed with some basic knowledge of dispositions. I do six things in this section as a means towards providing these arms.

First, I highlight the fact that our ordinary notion of dispositions provides a significant amount of familiarity with the concept. Second, I refine this ordinary notion by providing a more technical definition of the term based in the work of Neil Williams and Barry Smith. Third, I distinguish a handful of important kinds of dispositions, including *powers*, *capabilities*, *functions*, and *predispositions*. Fourth, I explain three features of dispositions that are central to my account: their *triggering conditions*, *manifestations*, and *realizations*. Fifth, I discuss what is sometimes called the ‘material base’ of dispositions. Finally, I distinguish and explain three different senses of the *reliability* of dispositions. This last task is key to one of the main elements of the dispositionalist account I defend: systematicity.

3.3.1 Disposition as a Familiar Notion

What is a disposition? Philosopher or not – indeed, metaphysician or not – my guess is that you already have an answer to this question. At the least, I suspect you have some general sense of what a disposition is. This is because, as we will see, these are quite familiar entities, and understanding them is a rather intuitive exercise. Their familiarity is actually helpful, too, since it means that we do not have to stray very far afield into the depth and complexities of modal metaphysics. Of course, we will do *some* modal metaphysics (including learning what ‘modal metaphysics’ means). But coupling the intuitive discernability of dispositions with a few neutrality-facilitating assumptions keeps us from having to run the modal metaphysics gamut.

That said, while the goal here is to lay out the basic elements of an intuitive, familiar property, it is nonetheless worthwhile making sure that we are on the same page. These basic elements serve to buttress the core of the positive account defended herein.

3.3.1.1 Ordinary Cases of Dispositions

Start by considering the following ordinary scenarios:

- (i) **Bubble Wrap:** Darling buys her sister Sarah an antique vase for her mantle. Darling knows that Sarah's two Siamese cats are quite mischievous, often climbing on furniture and knocking things over. This worries her, but she knows Sarah would love the vase and buys it anyways. Before leaving the store, she asks the clerk to wrap it in bubble wrap for the ride home.²⁰²
- (ii) **Easily Bruised:** Jo, a 70-year-old grandmother, has just arrived at her daughter's house for a holiday gathering where her young, rambunctious grandchildren are running about the house. Before she is helped inside, her daughter and son-in-law round up the grandkids and move their running and playing into the backyard. Jo is thankful and enters the house at ease.
- (iii) **Gimme Putt:** Adam, a professional golfer with 13 PGA Tour Wins, including a win at the prestigious Master's Championship, sets up to attempt an easy putt less than one foot from the hole. Adam misses the putt and is upset. As he misses, a TV commentator (also a professional golfer) rhetorically asks, "Oh my, what were you doing there?"²⁰³
- (iv) **Flying Guitarist:** Jimmy is a professional musician who can play many instruments, including the guitar. He is on a flight heading to visit his family for vacation. There are no instruments on the plane – his or anyone else's. He strikes up a conversation with the passenger seated next to him, who eventually asks, "Can you play the guitar?" Jimmy answers, "Yes, I can."

Consider some questions that are *not* very likely to come to mind about such commonplace occurrences. Why is Darling worried about the behavior of Sarah's cats? Why does she ask the clerk to bubble wrap the vase (and why do they, presumably, comply as if this is expected)? Why do Jo's kids move the children outside, and why is Jo relieved that they have done so? What makes Adam so upset and the commentator so surprised? Why do we find their respective attitudes seemingly justified?

²⁰² This case is inspired by characters from Disney's "Lady and the Tramp" (Luske et al., 1955).

²⁰³ This case is inspired by Austin (1956), as well as Adam Scott's actual missed putt at the 2017 U.S. Open. See the missed putt and hear the commentator's remarks here: https://www.youtube.com/watch?v=jBPqa_sbg48.

And why does Jimmy’s answer make so much sense despite his being stuck on a plane some 6 miles above any guitar? The answers to each of these questions come easily, and each is about dispositions.

3.3.1.2 Modal vs. Categorical Properties

Dispositions are *modal properties*. We can think of a property as, roughly, *a way that something is*. Sarah’s new vase is fragile. It is also, we might assume, blue with curved sides. These are ways that the vase is. Similarly, given her age, Jo is excessively sensitive to bruising. She is also 5 ft. 4 in. tall and has gray hair. Adam is an expert putter. Jimmy is a guitarist. These are all ways that the respective entities are – they are *properties* of these entities. Not all properties are the same, though. Some, like the dispositions we are interested in, are *modal*. Modality concerns necessity and possibility – it is about what *must* or *can* happen. Laws of nature are modal, for instance. If they are indeterministic, they dictate what *can* happen. If they are deterministic, they dictate what *must* happen.²⁰⁴ Dispositions are modal properties, then, because they are about what can or must happen. You might ask, “What can or must happen *with what?*” Dispositions are properties of the entities that bear them, and so being a modal property means that a disposition can be understood as *a way an entity is that explains what that entity can or must do*. The explaining invoked here can be understood as *truth-making*. Dispositions *make it true that* the entities bearing them can or must do something.²⁰⁵

Contrast modal properties, like dispositions, with what are sometimes called ‘categorical’ properties. Unlike modal properties, categorical properties do not point, as it were, towards possible or necessary events or processes that have yet to occur. As Williams puts it, “With categorical properties, what you see is what you get; there is nothing more to the property that is hidden.”²⁰⁶

²⁰⁴ Of course, what must happen is also something that *can* happen. That is, necessity, like actuality, entails possibility.

²⁰⁵ I mostly concentrate on possibility rather than necessity. Moreover, notice that this is still neutral concerning what dispositions are. For instance, even if one analyzes dispositions in terms of true counterfactual conditionals, it would still be the case that *the counterfactuals being true* explains – makes true – what the disposed entity can or must do.

²⁰⁶ Williams (2019, p. 29).

Properties commonly thought to be categorical (non-modal) are *shape*, *weight*, *color*, and *quantity*. Categorical properties, then, are not *poised* to do things – they are already being all they can be. Back to dispositions.

Consider *fragility*, a disposition of Sarah’s vase. It is the vase’s fragility that explains the fact that the vase will break (at least with some significant likelihood) if her cat knocks it to the hard floor. Jo’s being disposed to easily bruise (a disposition resembling, but different from, *hemophilia*) makes it true that, were her grandchild to bump into her, she would quickly form a serious bruise. Adam’s putting expertise makes it true that, roughly, he makes close-range attempted putts with some abnormally high regularity.²⁰⁷ This is precisely why Adam’s and the commentator’s responses seem appropriate to us – we are all justifiably surprised by his miss, likely thinking, “He should have made that!” Finally, like Adam’s putting expertise, Jimmy’s expertise at guitar is also a disposition. His disposition makes it true that, given the opportunity, Jimmy succeeds at playing guitar when he tries. That is, it is true now that Jimmy *can* play the guitar.

3.3.1.3 Two Things Dispositions Tell Us

Thus, dispositions tell us at least two things. First, they tell us about the present. They tell us that, right now, some entity is some way. Second, they tell us about some possible future. They tell us that, were certain conditions to be met, then some other possible process (involving the object bearing the disposition) can unfold. Neil Williams makes a similar point in his recent book on the metaphysics of modality:

When we say of an object that it has a power or a disposition, we seem to be saying a number of things. First of all, we are saying something about the object as it is *right now*, regardless of what else may come to pass. To say of the salt that it is soluble is to say something about the way the salt is, something about its nature.

²⁰⁷ The *abnormally high* regularity, or something like it, is needed to separate levels of putting skill. Regularity is likely present with any skill (a certain sort of disposition), and its levels correspond to levels of the skill. For instance, when the regularity is abnormally high, a *good* putter might become an *expert* putter.

Second, just as much as we are saying something about how the salt is right now, we are also saying something about how the salt could be in the *future*, or what it might do, should it find itself in the right (or wrong!) circumstances. It is a warning of sorts, an indication about specific ways in which the salt or the world around it is liable to change if the salt should find itself in some equally specific context. . . It is this second feature of powers that justifies our calling them *modal* properties, as their mode of presentation is not restricted to how they now appear.²⁰⁸

We will explore these additional components of dispositions in more detail below. For now, the upshot is that dispositions are undoubtedly familiar to us, at least in some basic sense. They include the *fragility* of glassware, the *solubility* of table salt, the *sharpness* of knives, the innumerable skills people have like *putting expertise* or *woodworking craftsmanship*, and various temperaments and personality traits like *friendliness* or *shyness*. While dispositions are familiar, we need a bit more precision about how exactly to understand them. The next subsection provides a start.

3.3.2 Defining ‘Disposition’

I begin by clearing some terminological ground so that our familiar understanding of dispositions can become a bit more refined. This will help to circumvent potential confusions as we unpack the nature of dispositions and then move into the details of the dispositionalist account of addiction in later sections.

Let us first start with a definition of ‘disposition’. Beyond my elucidation, I appeal here to definitions provided by Neil Williams and Barry Smith. While slightly different, these definitions are similar enough to be useful and appropriate for our purposes. Moreover, they are in line with the elucidation I gave above, that a disposition is *a way an entity is that explains what that entity can or must do*.

Here are two excerpts from Williams on how to understand dispositions:

Sometimes referred to as a ‘power’, or ‘propensity’, a disposition is the ability of an object to bring about some state of affairs (its ‘manifestation’), when met with the appropriate stimulus.²⁰⁹

²⁰⁸ Williams (2019, pp. 46–47).

²⁰⁹ Borghini & Williams (2008, p. 28). This understanding can also account for particles that, for instance, undergo spontaneous radioactive decay. The *trigger* – or, more accurately, *triggering circumstances* – here would simply be the state of affairs in which the particle is embedded. It is being in such circumstances that, together with the particle’s indeterministic (or stochastic) disposition, triggers the decay. Of course, it could have done so a moment before. This does not change anything. Some dispositions are probabilistic, and some are stochastic or random. This does not mean that the disposition

‘Disposition’, as I shall use the term, means nothing more than a capacity, power, or propensity of an object to act in some particular way in some particular set of conditions.²¹⁰

Consider next two definitions from Barry Smith, which are part of the Basic Formal Ontology (BFO). BFO contains a formal definition of ‘disposition’, and Smith and colleagues provide a natural language definition as well. They are as follows:

b is a disposition =Def. b is a realizable entity & b’s bearer is some material entity & b is such that if it ceases to exist, then its bearer is physically changed, & b’s realization occurs when and because this bearer is in some special physical circumstances, & this realization occurs in virtue of the bearer’s physical make-up.²¹¹

A *disposition* is a *realizable entity* in virtue of which — for example, through appropriate triggers — a process of a certain kind occurs (or can occur or is likely to occur) in the independent continuant in which the disposition inheres. This process is called the *realization* of the disposition... it is a realizable entity that exists because of certain features of the physical make-up of the independent continuant that is its bearer.²¹²

Again, while these definitions differ slightly in their descriptions, they are getting at the same basic idea I tried to capture above. First, dispositions are properties of entities; I said they are *ways an entity is*, Williams said they are *abilities or propensities of an object*, and Smith and colleagues said dispositions *have a bearer in which the disposition inheres*. This all comes to the same thing as far as we need be concerned.

Second, we each described these properties as *modal* in some way. I said they *explain what their bearers can or must do*. Williams and Smith add a bit more to their descriptions by specifying two important aspects of dispositions. The first is that there is some *particular set of conditions* or some *special*

does not require some set of circumstances it must be embedded in in order to be realized. It is just that the circumstances that work together with the disposition to produce its realization are often more like ordinary causal stimuli. The same is true of dispositions that are always manifesting. Perhaps some set of dispositions is responsible for *persistence* of objects over time, and these constantly manifest as the object persist. It will still be true that the dispositions work together with the circumstances in which they are embedded in order to be realized – that is, to produce the continued existence of its bearer. If one does not want to call this ‘causation’ or the conditions ‘triggers’ or ‘stimuli’, that is fine. I am not here trying to give an account of what causation is. Williams (2019, Sect. 7.4) provides a fuller discussion of the point I am trying to make here about dispositions being *triggered* (or *causal* for Williams) in cases of apparently missing stimuli and continually manifesting powers like gravitational force.

²¹⁰ Williams (2005, p. 304).

²¹¹ Buffalo Developers Group (2020, p. 11).

²¹² Arp et al. (2015, p. 101). Notice that this definition says the bearer is an **INDEPENDENT CONTINUANT** as opposed to the more restrictive **MATERIAL ENTITY**. This is probably preferable, since perhaps non-material entities like fields or God can bear dispositions. We can simply understand the first definition as getting at *material* dispositions.

physical circumstances that the bearer of the disposition must be in for the disposition to be *triggered*. The second is that there is *some state of affairs* or *manifestation* that the disposition can *bring about*, the bringing about of which is its *realization* (what it can or must do).²¹³ This, again, all comes to roughly the same basic idea. Dispositions, when in the appropriate triggering conditions, explain (together with those conditions) what further states of affairs or processes involving their bearers can be produced.

Third, Smith adds an additional component involving the physical makeup of the bearer of the disposition. In particular, he says that dispositions both exist (inhere in their bearers) and produce the manifestations they do at least partly because of the way their bearers are physically. In other words, for a vase to be fragile is, in part, for the vase to have some physical makeup (such as some type of molecular structure). Depending on the physical makeup of its bearer in this way entails that a disposition can disappear (or appear, if it was not there to begin with) if an entity physically changes in the right way. For instance, heating a glass vase in the right way can change its physical structure such that the disposition towards breaking is at least greatly weakened, and perhaps eliminated altogether (such that it would now be inaccurate to call the vase ‘fragile’). I will take this condition on board as well, accepting that part of what it is for an entity to be disposed (and part of what explains its manifestations) is that the entity has some relevant type of physical makeup.²¹⁴

²¹³ Two points of clarification, the latter of which gets slightly into some weeds. First, I say ‘can’ because not all dispositions will fire determinately when triggered (when in the appropriate circumstances). For instance, a vase may hit just right and not break when dropped in a typically-vase-breaking way. This also explains why Adam Scott can miss a 10-inch putt yet still be an expert putter. Dispositions explain what their bearers *can* do in the right circumstances. Manley & Wasserman (2007) and Vetter (2013, 2014, 2015, 2021) discuss degrees of dispositions. Second, Smith (and BFO) require that a disposition is realized in a process (like *the vase’s breaking*). Williams speaks of dispositions having “states of affairs” (like *the broken vase*) as their manifestations. So, one might think these descriptions are inconsistent. However, realizations and manifestations are not identical. A realization is *that which reveals a disposition for what it is* (what it can do), and so realizations are always changes (occurents). A manifestation is *the result of the realization of a disposition*, which can be either the change(s) realizing the disposition or the states of affairs they bring about. Hence, Williams’ and Smiths’ definitions are consistent.

²¹⁴ Note that this does not require that the physical makeup be so particular that, for instance, changing one atom destroys the disposition. Of course, this seems possible for some dispositions, but allowing this dependency between dispositions/manifestations and the bearer’s physical makeup certainly does not *entail* this kind of sensitivity. For instance, a disposition may depend on the bearer having some certain *type* of physical makeup (as opposed to some *particular token* physical makeup), and many possible specific physical makeups (or arrangements of fundamental physical stuff) could count as an instance of that *type*. For instance, there are likely many particular molecular arrangements that can make a vase fragile. Hence, this condition need not force us into thinking that dispositions are so sensitive as to be eliminated with any miniscule change to its bearer’s physical makeup.

Now that we have fleshed out and refined our central concept a bit more, it is worth saying something about the various terms that are used when people engage in ‘disposition’ talk. One common consequence of a concept being so ordinary is that it can be co-opted for various uses, which can thereby engender related terms and concepts (and the risk of confusion or misunderstanding). Sometimes these terms are synonyms, sometimes they are closely related but distinct, and sometimes it seems that they are both (or at least *used* in both ways). For instance, Williams uses the term ‘powers’ in the above passage. Are powers the same as dispositions? Moreover, Adam’s and Jimmy’s dispositions seem like *abilities*, but ‘ability’ seems like a less appropriate description of the vase’s fragility and of Jo’s disposition to easy bruising. In the next subsection, I address this concern.

3.3.3 Speaking of Dispositions: ‘Powers’, ‘Capabilities’, and Other Dispositionalist Terms

Using the foregoing refinement of our understanding of dispositions, we can now go through some commonly used dispositional terms. This will help to clarify for our purposes which are synonyms, which are not, and how they differ.

3.3.3.1 Two Points of Clarification

Let me first start with two points of clarification that will help us to understand the following discussion of other dispositional terms. First, dispositions can be understood as existing at various levels of reality. That is, electrons and cells have dispositions, as do organs (such as hearts and lungs), whole organisms, and perhaps even groups of agents like organizations. Anything that, because of the way it is, bears a property that explains what it is able to do can be said to have a disposition. Sometimes dispositions at higher levels are in some sense explained by dispositions at lower levels. For instance, my disposition to make tennis serves inheres in *me* – the whole person. But this ability surely rests on dispositions inhering in my eyes, fingers, muscles, and brain, among other parts of me. That is, it is my hand-eye-coordination, my muscle-memory, the grip of my hand that allows a consistently suitable

toss, and certain entrenched neural pathways processing all of this which partly make up my disposition to make tennis serves. These are or involve dispositions themselves, some of which may again rest on still lower-level dispositions (of cones and rods, muscle fibers, neurons, and so forth).

In addition, we can sometimes talk about this disposition hierarchy occurring at a single level, such as the level of the person. For instance, my hand-eye-coordination is plausibly a disposition inhering in *me*, the person (as opposed to just my eyes). It is a disposition *I* have which involves or rests on other dispositions of my eyes, hands, and so on. But my disposition to make tennis serves is not necessarily at some higher level despite resting on my hand-eye-coordination, among other things; both dispositions inhere in me, the person. Sometimes, to say that one disposition inheres in an entity means that some collection of (appropriately interrelated) dispositions inheres in that entity, and we give that collection (or collections like it) a particular name. Thus, we do not always need to go to some higher level of reality to find a disposition that rests on or is made up of other dispositions.²¹⁵

The second point of clarification is that I will almost always use ‘disposition’ more broadly than its related terms. Thus, other dispositionalist terms like ‘power’, ‘capability’, and the like will almost always refer to subtypes (or instances of subtypes) of the more general type **DISPOSITION**. Now to the related dispositionalist terms.

3.3.3.2 Powers: The Most Fundamental of Dispositions

I will start with the dispositionalist term ‘power’ (or ‘powers’). As we saw in the passage from Williams, this term is sometimes used interchangeably with ‘disposition’. To be sure, Williams himself

²¹⁵ A concerned reader might wonder whether I am committed to every dispositional ascription referring to some real, distinct property, or if we can be reductionists of a sort. In other words, does my disposition to make tennis serves *really* exist if it can be explained by a collection of other dispositions (and so on down the line)? As far as I can tell, this project does not require committing one way or the other on this question. Even the principle of *perspectivalism* adopted by Smith and the OBO Foundry only says that good ontologies must *treat* entities at each level of granularity *as if* they exist. To be sure, ontologies ought to be realist – true to reality. But there is also a practical use for ontologies, which does best when *perspectivalism* is adopted. Moreover, and perhaps most importantly, ontologies allow for *defined classes*, which entail less of an ontological commitment than do *universals*. Nothing in my project hangs on all dispositions, nor *addiction in particular*, being a universal or a defined class. We can still do all the work we need to do and get all the same results.

does not actually take these terms to be synonymous, and in the passage is only pointing out that common usage often does.²¹⁶ Indeed it does, as is seen in common phrases such as, “The President has the power to declare war” and “Maria has the power to fire Willie” – these likely refer to properties and relations constituting *authority*, and authority very likely is or involves dispositions. Closely resembling Williams’ usage, I will understand the term ‘power’ (and ‘powers’, ‘powerful property’, and so on) to refer to the most fundamental of properties that are also dispositional. Let me explain.

Recall the disposition hierarchy discussed above, according to which many dispositions will be made up of – or grounded in, or constituted by, or explained by, or whatever your preferred expression is here – further dispositions, which are sometimes more basic in some sense. In the example of my tennis serving ability, we saw that this need not always take us (at least not right away) to some lower level of reality. Still, this will typically be where we end up if we keep digging. Hence, by calling powers the ‘most fundamental’ dispositional properties, I mean to say that there is no more digging we can do that will turn up even more basic dispositions on which they rest (or are grounded in, or constituted by, or whatever you like).

As a kind of toy example to help illustrate the distinction, consider the difference between a dispositional property of a quark, such as its electric charge, and my ability to make tennis serves. Let us assume that quarks have no constituent parts, and so no part with its own more basic properties, dispositional or otherwise. Clearly my tennis serving ability rests on other dispositions (hand-eye coordination, learned muscle memory, and so on). The quark’s charge, we are supposing, does not. On the vocabulary I am proposing then, the quark’s electric charge would properly be called a ‘power’ since it is a disposition for which there is no more basic dispositions that might explain it (or ground

²¹⁶ See, for instance, Williams (2019, pp. 53–56).

it, or constitute it, ...). It is still a disposition, though, and therefore is an example of what I mean by a most fundamental dispositional property – a *power*.²¹⁷

3.3.3.3 Capabilities and Functions: Dispositions of Interest

Now let us move up, as it were, from powers to another kind of disposition, which I will call a ‘capability’. This terminology comes directly from an important ontology project that is currently under development: The Cognitive Process Ontology (CPO), initially developed by David Limbaugh, Eric Merrell, and Barry Smith, and which includes **CAPABILITY** as one of its classes.²¹⁸ The (Aristotelian) definition of ‘capability’ in CPO is as follows:

²¹⁷ Stepping slightly into the weeds a bit, two points ought to be kept in mind. First, my appeal to quarks and apparently fundamental, irreducible (in some sense) dispositional properties is not meant to provide a defense of *power monism* – according to which all properties are powers, at least at bottom – nor of *dualism* – according to which at least some fundamental properties are powers, while some are categorical. Some have made arguments like this, wherein the dispositions of things like quarks are meant to show that powerful properties exist which have no causal basis, and so which cannot be reduced to any further property, categorical or dispositional – they are “ungrounded” (Mumford, 2006). This so-called ‘argument from science’ has been contested even by friends of powers (Williams, 2009, 2011). For my part, the example is only meant to highlight the difference between dispositions which rest on (sometimes more basic) dispositions, and those which do not – *whether or not the latter have a causal basis in further categorical properties*. Again, I cannot see that anything in my project hangs on this debate.

Second, when Williams distinguishes dispositions from powers, his discussion of the former refers to *ascriptions* of dispositions. In this way, while we both differentiate the two based on some sort of appeal to degrees of fundamentality, Williams seems to take dispositions to be sensitive to usage (to actual dispositional ascriptions) in a way powers (and my use of ‘disposition’) are not – at least not always. Consider a case from Williams. Suppose an angel takes an interest in some fragile glass such that they always intervene just at the right moment, masking its fragility (and so not changing the glass at all) and always saving it from breaking when it otherwise would have. Suppose further that the glass’s owner, obviously surprised by this, spends months trying to break the glass, always failing (thanks to the angel). For Williams, the sense in which the glass is still fragile (despite the fact that it will never actually break) pertains to its powers: “...there are some properties possessed by the glass in virtue of which it counts as fragile, and the powers theories takes those properties to be powers” (2019, p. 54). However, Williams also contends that it makes sense for the owner (and others) to, at some point, stop *calling* the vase ‘fragile’ (a dispositional ascription), since they have discovered that it will almost certainly never behave in ways that typically justify such ascriptions: “The other perspective—that which sees the glass as non-fragile—belongs in the disposition ascription camp, as it concerns not just properties in the world, but also human elements, such as how we use terms like ‘fragile’...Unlike the former, the latter is highly sensitive to context, and possibly our interests too” (2019, pp. 54–55). So, while I am highly sympathetic to Williams’ reasoning here, for the purposes of my project I will nonetheless slightly depart from his terminology by taking ‘disposition’ to refer to high-level dispositional properties that are not merely constructed from the way we talk about things. In short, I think he is right that the glass is still fragile, but I think it is also fine to continue calling it ‘fragile’. This is *because* I agree with him about its fragility being unchanged. We have a perfectly good explanation for the lack of typical fragility behavior – the angel is masking its disposition(s). Perhaps it would be pragmatic to come up with another word (or call it ‘masked fragility’ or some such thing), so that we can differentiate it from glasses unprotected by angels. Nonetheless, the glass still seems just as disposed to break no matter what level of granularity we are at. Thanks to David Limbaugh for helpful discussion of these two points.

²¹⁸ CPO is an extension of BFO and some OBO Foundry ontologies, and therefore is being built to comply with the BFO’s realist methodology and the principles of best practice. See Limbaugh et al. (2020; 2019) for discussion.

CAPABILITY (A) is_a DISPOSITION (B) whose realization an organism or group of organisms has an interest in (that Cs).²¹⁹

While the definition is under development, we can go some way towards cashing this out a bit more. In doing so, I again follow the CPO developers' work. We already have an understanding of dispositions and a preliminary understanding of realizations, and so it is just the last notion about *organisms having an interest in realizations* that needs explaining. Limbaugh and company point to three principles that are meant to elucidate this notion.²²⁰

First, all organisms have an interest in the realization of the functions of their parts. In BFO, functions are a kind of capability, and hence a kind of disposition. Specifically, according to BFO, a function is, roughly, a disposition whose realization explains why the bearer of that disposition came to exist with the physical makeup it has.²²¹ For instance, if hearts have the disposition to pump blood throughout the body, and the realization of this disposition – the actual pumping of blood through the body – explains why the heart came into existence with the physical makeup it has, then the disposition to pump blood throughout the body is the function of the heart.²²² Artifacts provide another familiar example. If the realization of a hammer's disposition to force nails into substances (such as wood) explains why the hammer came into existence with the physical makeup it has, then the disposition to hammer in nails is the function of the hammer. The formula is, roughly: if *y*'s doing

²¹⁹ Limbaugh et al. (2019, p. 1). Note that *S has an interest in x* does not entail *x is in S's interest*. I might have an interest in going to the store to buy groceries even though, perhaps unbeknownst to me, this would overdraw my bank account. I might have an interest in using a hammer to drive in nails even though I am bad at it and it would likely smash my thumb if I tried. An addict might also have an interest in using heroin even though it is bad for them. Having an interest in *x* is not synonymous with *x* being good for you. Thus, this sense of 'interest in' is value-free in an important sense (though it *might* involve subjective value or evolutionary value). Thanks to David Limbaugh for his insight here.

²²⁰ See Limbaugh et al. (2019, pp. 2–3).

²²¹ Here is the more precise definition: "A **FUNCTION is_a DISPOSITION** that exists in virtue of the bearer's physical make-up and this physical make-up is something the bearer possesses because it came into being, either through evolution (in the case of natural biological entities) or through intentional design (in the case of artifacts), in order to realize processes of a certain sort" (Buffalo Developers Group, 2020, pp. 11–12). The font is modified to fit my typography.

²²² That might have sounded a bit odd. If not, please proceed. But if you thought I was going to say that the heart's function is *pumping blood throughout the body* (as opposed to being disposed to do so), then you would be (understandably) confusing function with functioning. The *functioning* of the heart is a process (such as the process of pumping blood throughout the body) while the *function* of the heart is a disposition which is realized in functionings. Thanks to Barry Smith's many post-presentation Q&A exchanges that allowed me to head off this confusion.

x explains why y exists (or was created) with the physical makeup it has (which grounds the disposition to do x), then being disposed to x is y 's function. From this, it follows that all functions are capabilities in BFO. For any biological function of a part of an organism, at least that organism will have an interest in its realization. Moreover, for any artifact function, at least the creator (and likely many actual and potential users of the artifact) will have an interest in its realization – in its *functioning*.

Limbaugh and colleagues' second principle elucidating the notion of *having an interest in disposition realizations* concerns plans, understood broadly as having an objective or goal. Even an intention counts as a plan in this sense, whether or not a plan is actually written down. The principle is that, for any plan some person or group has, that person or group will also have an interest in any process relevant to realizing that plan. Consider, for instance, processes that are *necessary* for a plan to be realized, which is a strong kind of relevance. If Ani plans to watch the new *Star Wars* film in 3D at the local theater at 5pm today, then Ani has an interest in making his way to the theater by 5pm and obtaining a pair of 3D glasses. Without either of these, Ani's plan cannot be realized. Hence, a disposition whose realization is relevant to – facilitates, is necessary for, is constitutive of – the realization of a plan is a capability. Ani's car's disposition to get him to the theater by 5pm is such a disposition, and so also a *capability* of his car.

The third principle concerns instrumentality, and is something of a transitivity principle. It says that, for any disposition x whose realization an organism has an interest in, and for any disposition y whose realization facilitates the realization of x , that organism will have an interest in the realization of y . Going back to Ani, this would entail that Ani, having an interest in the disposition of his car to get him to the theater on time, will also have an interest in the realization of the engine's disposition to turn over, the realization of the radiator's disposition to cool the engine, and so on. This is because such realizations facilitate the realization of the original disposition – to get him to the theater on time. Thus, each of these dispositions would be capabilities.

To take stock, a capability is a disposition of some entity. It is a disposition just like fragility is a disposition, at least in the sense that they both share the features of dispositions discussed above (being modal, being grounded in the physical makeup of their bearer, being realized in processes, and so on). Capabilities, though, are such that some organism or group of organisms has an interest in their realizations. As with Ani's car and the hammer, this need not be the bearer of that disposition.

3.3.3.4 Predispositions: Dispositions to Dispositions

Consider one more kind of disposition before moving to a discussion of some of the other features of dispositions. These are *predispositions*, and they are really quite simple. A predisposition is *a disposition to acquire some other disposition*. For instance, often scientists will speak of 'genetic predispositions' for certain diseases, and such claims have even been made regarding addiction.²²³ Whether or not there is such a genetic predisposition, the idea is simply that an organism, in virtue of their genetic makeup, is disposed towards acquiring some other disposition, which is itself a disease (a disposition towards certain pathological processes²²⁴).

Another familiar example is my predisposition to speak German. I cannot, right now, speak German. Of course, there is some sense in which it is true, right now, that I *could* speak German. I am *now* disposed to acquire other dispositions (through finding and taking German lessons, for instance), which would then more directly dispose me towards speaking German. This sense in which I am disposed towards *being able to* speak German (as opposed to being disposed towards *speaking German*) is my predisposition to speaking German.

²²³ Some experimental studies, for instance, assume there is a genetic predisposition for certain addictions (like alcoholism) and then study how this can be influenced or counteracted (Froehlich et al., 2017). Some have even suggested there are genetic predispositions for *resisting* addiction since they are related to self-control (Kendler & Myers, 2015). The scientific community is far from a consensus on whether there is some gene or group of genes responsible for predisposing individuals to addiction. This almost certainly has to do with the numerous complexities involved in such a claim. For instance, a predisposition towards impulsiveness, persuadability, or risky behavior (which are insufficient for addiction) must be teased apart from a predisposition to addiction *per se*.

²²⁴ This understanding of 'disease' is used in the Disease Ontology and the Ontology for General Medical Science, which extend BFO (Ceusters & Smith, 2010; Cowell & Smith, 2010; Scheuermann et al., 2009).

One worry with introducing predispositions is that we might produce an explosion of dispositions. Am I predisposed to engage in karate on Mars because I am some way now such that, through a series of realizations of dispositions, I *could* be on Mars knowing karate? I have two responses. First, maybe I am predisposed in just this way. After all, even the more intuitive cases of predispositions like genetic predispositions and predispositions to learning languages are not separated from their more direct dispositional counterparts (having a disease, knowing German) by only a single manifestation. I do not learn German by some predisposition I have manifesting in a single process, and then, *voilà!* Or, perhaps, *hier hast du es!* Moreover, we need not say that I am *disposed* towards engaging in karate on Mars. We are saying that I am now such that, were a certain (long) series of (admittedly improbable) realizations of dispositions to occur, *then* I would be disposed towards doing karate on Mars. I'm not sure this is such a terrible thing to accept.

Second, and less conciliatory, it is possible to go some way towards avoiding this explosion, or at least containing it. We could say that predispositions, while not typically connected to their acquirable dispositional counterparts through a single manifestation, still require some degree of what we might call 'modal proximity' between them. In other words, it matters how much has to change (about you, about the world) in order to get from one disposition to the other. There may not be a hard and fast rule, but if a (purported) predisposition is too *modally distant* from its (purportedly) acquirable dispositional counterpart, then it may be no predisposition at all. At least, it would not be one we would care about.²²⁵

One final note. There are other dispositional terms (perhaps many), such as 'capacity', 'ability', 'tendency', 'temperament', 'inclination', 'propensity', 'proclivity', 'proneness', 'predilection', and so on.

²²⁵ If one is inclined to have a name for modally distant predispositions, then 'pre-predisposition' comes to mind (though, so does a looming regress of 'pre'-s). In the end, though, nothing hangs on what we call them.

For my purposes, I will consider all of these to be either synonyms of ‘disposition’ in the more general sense, or capable of being differentiated in some relevant but yet-to-be-determined way.

3.3.4 Dispositions *for* Something: Triggering Conditions, Realizations, and Manifestations

We have been introduced to the familiar notion of dispositions and have seen some more refined definitions of both ‘disposition’ and related dispositionalist terms. Before filling out the dispositionalist account of addiction in Section 3.4, it will be useful to elaborate further on five important additional features of dispositions. We will start with triggering conditions, realizations, and manifestations in this subsection. The following two subsections will cover material bases and what I call the ‘reliability’ of dispositions.

All dispositions have *triggering conditions*, which are simply those (types of) circumstances in which a disposition *can* be realized when it is embedded in them.²²⁶ A *realization* of a disposition is, roughly, the changes (processes) which reveal the disposition for what it is. When a disposition is realized, it produces its *manifestations* – those states of affairs or processes the disposition *can* produce when triggered (either its realization or the results of its realization).²²⁷ For instance, being dropped onto a hard floor is a *triggering condition* of fragility, the breaking of the fragile object is a *realization* of fragility, and both the breaking and the subsequent broken object are *manifestations* of fragility. Dispositions are always *for* some type(s) of process, its realization(s). This is what is meant by the fragile vase being *disposed towards breaking*. Moreover, the triggering conditions help bring about what the disposition is for, as when we say that the fragile vase is disposed towards breaking *when suitably struck*. In this way, triggering conditions can also be understood as realization (and manifestation) conditions. A disposition is always a disposition *towards x when y*.

²²⁶ The realization is produced (or occurs) because of the disposition *together with* the triggering conditions.

²²⁷ Thus, when a dropped vase’s broken pieces cut my foot, this is not a *realization* of fragility since that process itself does not reveal the fragility (as opposed to the *sharpness* of the pieces) for what it is. However, it would strictly-speaking be a *manifestation* of fragility since it is a process that, in virtue of being fragile, the vase can produce if appropriately triggered.

Two questions arise. First, can dispositions have multiple types of triggering conditions (and realizations and manifestations)? Second, who decides which count?

3.3.4.1 Multi-Track Dispositions

To the first question of whether dispositions can have multiple types of triggering conditions, realizations, or manifestations, the answer is that they can. Regarding triggering conditions, it is common to try to capture the many different kinds of circumstances that can trigger a disposition by describing them at a high enough level. For instance, many different types of circumstance might trigger fragility (dropping, kicking, hitting with a bat, squeezing, and so on). However, we might try to capture all of these with a single type of event: *being suitably struck*. We should keep in mind that, while sometimes appropriate, this is often just a heuristic. Consider that a fragile glass may break due to a high-pitched sound or by increasing the gravitational force to some degree. Neither of these are striking. The heuristic can be convenient, though, and in discussing addiction I will make use of it.²²⁸

Regarding dispositions with multiple realizations and manifestations, Neil Williams (following Gilbert Ryle) calls these ‘multi-track’.²²⁹ Multi-track dispositions can be realized in different ways, bringing about different kinds of manifestations. Consider the property of **BEING LIQUID**, which we can assume for the moment is a kind of disposition. Liquids – in virtue of being liquid – are disposed towards *maintaining their volume, conforming to the shape of their containers, evaporating, boiling*, and perhaps other processes. Hence, it is no problem that some dispositions may have multiple types of circumstances that can trigger them, as well as multiple types of processes that count as their realizations (or multiple states of affairs that count as their manifestations).

²²⁸ Barbara Vetter discusses whether the nature of a disposition, such as fragility, is best understood in terms of a single conditional, such as *if x is suitably struck, then x will break*, or even any conditionals at all (Vetter, 2013, 2014).

²²⁹ He focuses on manifestations rather than realizations (Williams, 2019, Ch. 4, Sect. 2). However, realizations are a kind of manifestation on my view, and so his discussion is consistent with what I say here.

3.3.4.2 Direct vs. Indirect Manifestations and the Standard of Reality

Turning to the second question of who decides which circumstances count as triggers, realizations, and manifestations, the answer is simply that no one does. At least, no one decides this in the sense that a circumstance or process can become a triggering condition, realization, or manifestation of a disposition *simply by someone's deciding or declaring it to be so*. Turn again to the dropped vase that cuts my foot. It seems harmless enough to regard this as a manifestation of the vase's fragility since it is a process that the disposition can produce (indeed, it just produced it). But what about my subsequent crying, or the crying of the vase's owner (who loved the vase), or me bandaging my foot and driving to the hospital, or the owner's deciding to replace the vase? Is every link in this chain of events a manifestation of fragility? It might seem so since it seems true to say that the vase's fragility can bring them about. I have a few things to say about this.

First, we can distinguish between *direct* and *indirect* manifestations of a disposition. The difference roughly comes to what the disposition can *immediately produce*. This means that when a disposition is in the appropriate triggering conditions, direct manifestations can occur without any intermediate steps. Realizations are direct manifestations in this sense. Once suitably struck, for instance, a fragile vase can begin the breaking process (its realization). All other states of affairs that can be produced by a disposition would be indirect manifestations.

Second, we might add to the class of direct manifestations those that are the immediate result of realizations. For instance, the broken vase which is the immediate result of the realization of the vase's fragility (the breaking process) would be a direct manifestation in this sense. Similarly, *dissolving in water* is a realization of a portion of salt's solubility, and *the water's increased salinity* is an immediate result of its realization; both would be direct manifestations of the salt's solubility in this sense.

Third, we can distinguish indirect manifestations by invoking the notion of *modal proximity* introduced in our discussion of predispositions above. In other words, some indirect manifestations

are going to be more or less modally distant from the disposition and its realization. Moreover, there might be some relevant degree of modal distance such that processes or states of affairs that exceed it are no longer appropriately called ‘manifestations’ at all, indirect or otherwise.

These ideas are reflected in the BFO definition of ‘realization’ (modified for readability):

P is realization_of D means: there is some **MATERIAL ENTITY** or **IMMATERIAL ENTITY** E, and D is a **REALIZABLE ENTITY** that *inheres_in* E, and for all t, if P *has_participant* E at t, then D exists at t and the type instantiated by P is correlated with the type instantiated by D.²³⁰

The underlined clause expresses the idea that disposition types are, in some sense, paired with their realization types (or matched in the case of multi-track dispositions). This suggests a way to demarcate which processes count as realizations of a given disposition and which do not. BFO does not define ‘manifestation’ (or ‘realization’), but the paragraph above provides a way for manifestations in my sense to be distinguished from related states or processes that are not manifestations of some disposition. Moreover, the BFO method for distinguishing realizations from non-realizations aligns with my answer to the second question we were answering: who decides which states or processes count as realizations and manifestations? As the BFO definition suggests, reality does.

Dispositions have realizations and manifestations, and we can do our best to discover what those are (for example, through scientific and philosophical investigation). Fragility seems paired with breaking, for instance. Perhaps it has other realizations. Perhaps the vase owner’s crying *is* one of its manifestations. Either way, whatever is *in fact* the case is up to reality, not to any of us to decide.

3.3.5 Dispositions are Internally Grounded in their Bearer’s Physical Makeup

In addition to triggering conditions, realizations, and manifestations, dispositions have a material basis. I will follow BFO in its understanding of what it is for a disposition to have a material basis, which is as follows (slightly modified for readability):

²³⁰ Buffalo Developers Group (2020, pp. 10–11), underline added and font of class names modified to fit my typography.

D *has_material_basis* B at t means: D is a **DISPOSITION** and B is a **MATERIAL ENTITY**, and there is some E that is *bearer_of* D, and B is *continuant_part_of* E at t, and E is *bearer_of* D because B is *continuant_part_of* E at t.²³¹

Consider again my tennis serving ability as an example. Some of my parts, such as my eyes, my muscles, and parts of my brain, serve as the material basis of this disposition. This is because my having those particular parts, including the ways that they are, is the reason I bear the disposition that is *my tennis serving ability*. Moreover, losing (or initially gaining) the disposition requires some change in the material basis. If I were to go blind, for example, I would likely lose my tennis serving ability. However, it seems possible that I could get it back with a lot of practice.²³²

This last point highlights the fact that a disposition does not necessarily require some specific material basis. This is particularly true with high-level or complex dispositions like skills, character traits, and addiction. This is because dispositions like these can come in degrees in a certain sense. For instance, the strength, hand-eye coordination, and muscle memory of those with a *tennis serving ability* can each vary to some degree. One person may be slightly less strong but have better hand-eye coordination, and so on for the various other aspects of having this skill. Changes in such things as strength, hand-eye-coordination, and so forth will correspond with changes to the physical parts of the person (muscle fibers, brain states and pathways, and so on). This explains why each instance of a disposition need not have qualitatively identical material bases. The question of who decides might crop up again at this point. The answer, though, is the same: reality does.

3.3.6 Three Senses of Reliability: Strength, Opportunity, and Systematicity

The final point to consider is that all dispositions can be said to have what I call ‘reliability’. But there are different senses of a disposition’s reliability. Let us consider three that I take to be central.

²³¹ Buffalo Developers Group (2020, p. 12). Nothing about this definition rules out the possibility that two distinct dispositions can share a material basis.

²³² Supposing we are just talking about the *making* of tennis serves (given attempts in normal conditions), and not any subsequent winning of points. The latter would be much harder were I blind.

3.3.6.1 Reliability as Strength: The Likelihood of Being Realized when Triggered

The first kind of reliability pertains to a disposition's level of strength (or weakness), which is the likelihood (or probability) that the disposition would be realized when triggered.²³³ A disposition could be *fully* reliable in this sense, where this means it would be realized deterministically – that is, *without fail* – when triggered. For instance, suppose a set of deterministic laws entails a deterministic relationship between some disposition (together with its triggering conditions) and its realization. This disposition could not fail to be realized were it to be triggered, and thus it would be fully reliable in the sense of being *maximally strong*.²³⁴

However, many (perhaps most) dispositions will possess some non-maximal degree of strength.²³⁵ That is, when triggered, they will be realized only with some degree of likelihood. Skills are a good example of this. My tennis serving ability is by no means fully reliable. Let us suppose that, when attempted in normal conditions, I can make tennis serves with 85% likelihood. This means that even when I am in normal triggering conditions, and have the tennis serving ability, I will not necessarily make an attempted serve. Dispositions like this explain why their bearers can accurately be said to have the disposition despite their not always showing up when we might expect. When triggered, dispositions fire *according to their nature*, which is not always at maximal strength.²³⁶

3.3.6.2 Reliability as Opportunity: The Commonality of Triggering Conditions

A second sense of reliability pertains to the likelihood with which a disposition *can be expected to be triggered*. This pertains to how common (or rare) the disposition's triggering conditions are. In addition to their prevalence, it pertains to how likely it is for the relevant disposition(s) (or, more

²³³ BFO accommodates varying strengths of dispositions in basically this sense (Arp et al., 2015, pp. 101–102).

²³⁴ Smith and colleagues call these 'sure-fire dispositions' (Arp et al., 2015, p. 102).

²³⁵ See Manley & Wasserman (2007) and Vetter (2013, 2014, 2015, 2021) for discussions of the degrees and gradeability of dispositions.

²³⁶ Difficult questions arise as to how reliable the relationship between a disposition (paired with its triggering conditions) and its realization must be in order to count as a certain disposition. I try to say something about this below.

precisely, their bearers) to actually encounter their triggering conditions. This gets at a distinct sense of reliability since it is possible to have a maximally strong disposition that has extremely rare triggering conditions or a very weak disposition that has incredibly common triggering conditions. An example of the former is your disposition to suffocate in the stratosphere (or freeze in even higher layers of the atmosphere). While this disposition is likely very strong, there is only a very minimal likelihood for it to be triggered. An example of the latter is the disposition of a person who lives mostly indoors with all smokers to develop lung cancer from second-hand smoke. While this disposition is likely rather weak, there is quite a high likelihood that it will be in triggering conditions. Thus, reliability as mere opportunity is very much about external conditions. In contrast to a disposition's strength, nothing about the disposition or its material basis needs to change in order for it to be more or less reliable in this sense; it has only to do with *actual opportunity* of being triggered.²³⁷

3.3.6.3 Reliability as Systematicity: Sufficiently Strong in the Right Triggering Conditions

A third sense of reliability is what I call 'systematicity'.²³⁸ Systematicity pertains to a disposition's strength, but it is not *just* about strength. Systematicity also pertains to something analogous to, but distinct from, the opportunity sense of reliability. It is not just the combination of strength and opportunity. Let me explain.

To see why systematicity is distinct from strength as I understand them, consider the example of our disposition to suffocate in the upper layers of the atmosphere. This is an incredibly strong (probably nearly deterministic) disposition. However, we are virtually never in – nor likely ever to be in – the relevant triggering conditions. But this does *not* entail that there is no such disposition (bubble wrapped vases in vaults are still fragile). However, it suggests that this disposition is not a *systematic*

²³⁷ I do not think opportunity matters much in determining *that* a disposition is present. Like Williams, I still think the vase is fragile when we bubble wrap it and put it into the closet, never to be touched again.

²³⁸ Thanks to David Limbaugh for many helpful discussions of systematicity, and for letting me steal the name. See Limbaugh (2019) for his discussion of systematicity in the context of dysfunctions and disorders.

disposition to suffocate in upper atmospheric layers. Of course, if by ‘systematic’ you have the first sense of reliability in mind (strength, or reliability of firing when triggered), then it would be. But this is not what I mean by ‘systematic’ – I called *that* sense of reliability ‘strength’. Thus, we must be careful because it might sometimes feel tempting to say a disposition is systematic when it is reliably realized whenever triggered. It does not ultimately matter what names we use, but we need only mind our terminology so that we do not end up misunderstanding or talking past one another.

While distinct, strength does still matter to systematicity in that it is only when a disposition is sufficiently strong – realizing with some sufficiently high probability when triggered – in the right types of circumstances that it can be systematic on my view. That is, strength (with respect to certain triggering conditions) is a necessary condition of systematicity.

In addition, systematicity involves something analogous to the opportunity sense of reliability, but it is not simply about how common the disposition’s triggering conditions actually are. If it were, then it would just be about opportunity. Instead, in addition to sufficient strength in the right kinds of circumstances, systematicity pertains to which triggering conditions count towards determining whether a disposition that is sufficiently strong is also systematic. And this is where something analogous to, but distinct from, opportunity comes in. Systematicity pertains to whether a disposition is sufficiently strong in a sufficient number of triggering conditions *that members of the addict’s reference class are sufficiently likely to be in*. Let me spell this out a little more.

Consider *fragility*, which seems to be about breaking when dropped. Suppose we are trying to figure out whether bowling balls are fragile. Bowling balls are very strongly disposed to break when dropped from very tall buildings. Moreover (setting aside legal repercussions), there is plenty of opportunity for people to put bowling balls into these conditions if they want. Nonetheless, these two facts do not make bowling balls fragile. That is, they do not make it true that a bowling ball’s

disposition to break when dropped is *systematic*, even though bowling balls have a disposition, which is not fragility, to reliably break *when dropped from very tall buildings*.

Despite the available opportunities to drop bowling balls from very tall buildings, this is simply not a common circumstance for a bowling ball to be in since any given bowling ball has a rather low probability of being tossed from a tall building. This is likely true for most kinds of ordinary objects (vases, picture frames, chairs, and so on). As such, this is not the relevant circumstance for testing whether a bowling ball is fragile. But it is not because *this kind of triggering condition is rare* – indeed, it was noted that there *is* plenty of opportunity to put bowling balls in these circumstances. Instead, it is because *members of the bowling ball's reference class being in such a triggering condition* is rare. In other words, the claim that ‘bowling balls have a systematic disposition to break when dropped’ is not false because there is low opportunity for bowling balls to be dropped from tall buildings. Instead, because it is so uncommon for bowling balls (and most everyday objects) to be dropped from tall buildings, these conditions do not count in determining whether ‘bowling balls have a systematic disposition to break when dropped’ is false. If we lived in Skyscraper World, where everyone and everything was located on top of very tall buildings, then perhaps bowling balls would be fragile. Their strong disposition to break when dropped from tall buildings might be coupled with the fact that, for a given bowling ball (or any ordinary object), there is some significant probability that it will fall from a tall building. Of course, on this view of fragility as *the systematic disposition to break when dropped*, everything might be fragile in the context of Skyscraper World.²³⁹

The upshot here is that systematicity pertains to delineating which triggering conditions count towards some disposition being systematic. As we saw, this is about appealing to the appropriate reference class for the bearer of the disposition (such as all bowling balls or all everyday objects) and

²³⁹ Alternatively, we might change our understanding of what fragility is, or cease to care about it in the same way, or start caring about another disposition, *fragility**, whose triggering conditions are restricted to falls that do not include going over the edge of the tall buildings we lived on.

determining what the likelihood is for its members to be in those conditions. If it is sufficiently high, then those conditions can count towards the disposition being systematic.²⁴⁰ Expanding this further, the idea is this: for any conditions which can trigger the disposition and which members of the relevant reference class have a sufficient probability of being in, given the laws of nature and history of the world, a disposition is systematic when it is strongly realizable in sufficiently many of those conditions.²⁴¹ Keep in mind that this also means that *being dropped from very tall buildings* is also not going to be a relevant test case for the fragility of vases any more than bowling balls – vases are no more likely to be in such conditions. This is the right result, though, since being dropped from very tall buildings is not a very good test of whether something is fragile.

Finally, systematicity can help us make sense of contextualizing certain dispositions, where this means determining whether a systematic disposition is present *according to some context*. For instance, we might determine that Sam’s disposition towards violence is systematic *in prison* since that disposition is sufficiently strong in sufficiently many triggering conditions that *prison inmates* have a high probability of being in. However, the disposition may be much weaker in triggering conditions that *the general public* has a high probability of being in. It is in this sense that Sam can be described as ‘violent in prison, but not generally violent’. The same could be true of an addiction disposition.

3.3.7 Taking Stock of the Dispositional Features

In the next section, we will see what the dispositional account of addiction I defend looks like by adding some particular content to the framework of dispositions that has just been laid out. At this point, though, it might be useful to provide a sort of inventory for the relevant features discussed above. Here are some of the features of dispositions we have been introduced to:

²⁴⁰ I am neutral about whether we should reserve the term ‘triggering conditions’ for those that count or, instead, maintain the use of ‘triggering conditions’ for those circumstances that, together with the disposition, *can* produce its realization, and simply add the qualifier ‘relevant’ to those triggering conditions that count towards systematicity.

²⁴¹ This is an adaptation of the use of ‘systematicity’ in Limbaugh (2019).

- (i) A disposition is *a way that something is which explains what that thing can or must do*;
- (ii) Dispositions explain what their bearers can do in the sense of *making it true* that they can do that thing;
- (iii) Dispositions are *realized in* (or are *for*) certain processes (their *realizations*), which reveal the disposition for what it is;
- (iv) The states and processes that the disposition can produce, including its realization, are *manifestations* of that disposition;
- (v) Dispositions have certain *triggering conditions*, which are those circumstances in which a disposition *can* be realized or manifested;
- (vi) Dispositions have a *material basis*, making them internally grounded in their bearers (changes in the disposition entail changes in the physical makeup of the bearer);
- (vii) Dispositions have varying degrees of *reliability* in at least three senses:
 - a) **Strength:** the likelihood with which a disposition would be realized when triggered (those firing deterministically when triggered are *maximally strong*);
 - b) **Opportunity:** the availability or prevalence of a disposition's triggering conditions;
 - c) **Systematicity:** when a disposition is sufficiently strong in sufficiently many of those conditions which members of the bearer's reference class have a sufficiently high probability of being in, given the laws of nature and history of the world.

Let us now turn to filling in the content of the dispositional account of addiction.

3.4 The Addiction Disposition: Desires, Impaired Control, and Systematicity

Addiction is a certain type of disposition. With our foray into the world of dispositions behind us, this claim should be much clearer. If an *addict* is an organism that bears the addiction disposition, then addiction is *a way that organism is* which makes it true that, in virtue of having that disposition, the organism is liable to behave in certain ways under certain conditions. Moreover, an addiction might not be realized even when appropriately triggered since it is very plausibly a non-deterministic disposition.²⁴² The physical makeup of the addict – some arrangement of their parts and properties –

²⁴² Despite its *prima facie* plausibility (even its obviousness), discussions of compulsion in addiction often imply that addiction *would* be maximally strong, firing deterministically when triggered. This is seen, for instance, in many *defenses* of the brain disease model of addiction, wherein it is argued that addicts are irresistibly compelled and unable to do other than succumb to their desires. But this idea is also seen in many *criticisms* of such views, as when choice model theorists implicitly agree that, were addiction a disease, addicts would be irresistibly compelled. They then conclude with the claim that, since addicts are *not* compelled in this way, addiction is not a disease. While the former put the deterministic realizability into addiction more generally, the latter only see *addiction-as-disease* as entailing deterministic realizability. Either

will serve as the material basis grounding the addiction. There will be paradigmatic triggering conditions, realizations, and manifestations of addiction. Perhaps most importantly, addiction will involve systematicity – an addiction is present when the disposition to fail to control one’s desires towards certain types of behaviors is sufficiently strong in a sufficient number of triggering conditions that members of the bearer’s reference class are sufficiently likely to be in.

Recall that dispositions are *for* certain types of processes – fragility is *for* breaking (when suitably struck), putting expertise is *for* sinking putts (when suitably attempted), and so on for all dispositions. Hence, addiction must also be for some particular type(s) of process(es). This, I will argue, is at least *the failure to control one’s desires to engage in certain types of behavior.*²⁴³ Of course, bowling balls and tall buildings taught us that being at all disposed towards some type of process (such as *breaking* or *failing to control one’s desires*) is insufficient by itself to establish that some systematic disposition (like *fragility* or *addiction*) is present. This is true *even if that disposition is for processes of the type that occurred.* Put simply, this is because there may be a disposition present, just one that is not systematic. Hence, merely being disposed to fail to control one’s desires (to any degree whatsoever) is insufficient for addiction, just as merely being disposed to break is insufficient for fragility – it matters *how* disposed the bearer is.

A bowling ball’s disposition to break is not fragility, even if we can mimic the realization of fragility by dropping the bowling ball from the Empire State Building. Similarly, a person trapped in a room with only cocaine-laced water may eventually realize *some* disposition to fail to control their desires (which are perhaps newly acquired). But this need not imply an addiction, which is a *systematic*

way, both sides often imply maximal strength where they should not. Neither addiction nor disease entails deterministic realizations when in triggering conditions. If this sounds odd or implausible, welcome to the addiction literature.

²⁴³ Let me head off a concern about absences. The failure to control one’s addicted desires is not necessarily an absence on my view. This partly has to do with what I will go on to say are the triggering conditions, which are themselves positively occurring events, namely, *attempts* of a certain sort. Of course, we can describe *some* failures to control desires as absences, such as when I fail to control my desire for ice cream by being tied up (and so doing nothing), or by remaining seated on the couch, and so on. But I am not committed to these counting. Consider my putting skills, which are not even close to counting as expert. When I try (in suitable conditions), I am systematically disposed to miss – that is, to *fail* to make – attempted putts. But these are poorly executed putts, not absences. If you are averse to dispositions being realized in absences, then think of ‘failures to control one’s desires’ as referring to cases like my disposition to fail to make putts.

disposition to fail to control one's desires, is present. That situation is too much like the Empire State Building scenario to be conclusive. Of course, it is *possible* that this is a realization of addiction. But it is also possible that *anyone* in such circumstances would fail to control their desires, addiction or not, just as most things will break when dropped from the Empire State Building, fragile or not. The difference, I will argue, comes to systematicity. Before we get more into systematicity in addiction, though, I need to say more about the process that realizes addiction. Specifically, I provide some reasons for including its main components: desires and impaired control. I start with desires.

3.4.1 Desires as a Component of Addiction

As an initial motivation, consider that no author I am aware of who studies or writes about addiction denies the role of desires. Whether it is thought that desires compel addicts, are non-compelling but disordered, are non-disordered but abnormally difficult to resist, are just like those that everyone experiences in cases of *akrasia* (weakness of will), or are ordinary and resistible and hence immoral and condemnable, everyone agrees that desires have a role to play.²⁴⁴ Consider further the uncountable self-reports from addicts describing their experiences of strong urges, cravings, and other species of desire, as well as the widely acknowledged role of the desire- and motivation-producing mesolimbic dopamine system in addiction.²⁴⁵ To be sure, this does not imply that everyone agrees on the role of desires, nor on what desires themselves are. Nonetheless, this is quite telling, and in my view serves as a solid starting point for the claim that addiction involves, at least in part, desires (and *desirings*). We just need to say what desires are, and what their role is in addiction.

²⁴⁴ For such disparate views, see, respectively: Volkow & Fowler (2000); Heyman (2009); Lewis (2015) and Sinnott-Armstrong & Pickard (2013); Heather (2017a) and Foddy & Savulescu (2010a); and Peele (1987).

²⁴⁵ The role of the dopamine system is well-known (Kalivas & Volkow, 2005; N. Volkow, 2014; N. Volkow & Morales, 2015). See Lewis (2015) for a view from a neuroscientist who denies that addiction is a disease while accepting all of the relevant neuroscientific data regarding the involvement (and changes to) the mesolimbic dopamine system.

3.4.1.1 What Desires Are

Ultimately, my goal is to offer a framework for thinking about addiction that can accommodate different theories about the nature of some of the various components of addiction, such as desires and control. However, it will still be useful to have an account on hand. To that end, I will mostly (with some minor deviations) follow Nomy Arpaly and Timothy Schroeder in characterizing desires.²⁴⁶

Desires influence our actions, feelings, and thoughts. If we know that Andrew desires to go on a run every morning at 5 a.m., then we should be unsurprised to find him out running at 5:05 a.m. tomorrow. We should also be unsurprised to hear him tell us that he felt disappointed to have missed his run yesterday, that he felt satisfied after today's run and pleased to be back on track, and that he feels the urge to run as soon as he is awake and putting his running shoes on at 4:45 a.m. Moreover, we should expect Andrew to think about his morning runs in particular ways, given his desires. For instance, if Kristan invites Andrew to the gym this evening, he will likely think about how this will affect his morning run. If she invites him on a week-long trip to the beach, he will likely think about whether his morning runs can continue on the trip. He may even look up the hotel and city streets online so as to make plans for when and where he might run during the trip. The same is true of desires for chocolates, for peace and quiet, for finishing your paper, for seeing the Grand Canyon before you die, for doing the right thing, and for any other desire. This is just part of what desires do.

In my view, desires should be understood dispositionally. They are what dispose someone like Andrew to act, feel, and think in certain ways in response to representing the world to be some way (or to have been some way, or to likely be some way soon). But what is it that disposes us in this way?

²⁴⁶ See Arpaly & Schroeder (2014). I am also grateful to David Limbaugh and Eric Merrell for many helpful discussions about desires. Much of the following is greatly influenced by all three sources.

Arpaly & Schroeder (2014) provide an answer, and it is grounded in reward-and-punishment-based learning. Hence, they call their view the “reward theory of desire.”²⁴⁷

3.4.1.1.1 Reward Learning and Desires

We will skip over much of the details here, but the basic idea behind reward learning is as follows.²⁴⁸ Throughout our waking lives there is a constant flow of one mental state playing a causal role in bringing about a subsequent mental state.²⁴⁹ You perceive a chocolate on the table, and this causes you to think of its (potentially) delicious taste. This may cause you to feel presently hungry for chocolate, and this may cause you to decide to eat it. You hear a car door close just outside, and this causes you to form the belief that your roommate is back from the store. You see your running shoes, and this causes you to feel an urge to go running. This is happening all the time and with each causal sequence there is an opportunity for reward learning. That is, very often after the causal sequence of one mental state causally contributing to the formation of another, we receive feedback in the form of an unconscious learning signal. This tells us, roughly, whether how the world is now is better than expected, just as expected, or worse than expected. These positive, neutral, and negative learning signals, respectively, then either strengthen, do nothing to, or weaken the disposition of the first mental state to bring about the second in similar circumstances. So, if you see and then decide to eat the chocolate, and this is followed by a positive learning signal (“Hey, that was much better than expected!”), your disposition to decide to eat chocolates upon seeing them in similar circumstances has just been strengthened to some degree. A rough illustration of this is seen in Figure 10 below.

²⁴⁷ I will mostly just speak of ‘reward learning’ and will say something below about how punishments relate to learning and desires (or rather, aversions). It will roughly be the mirror image of reward.

²⁴⁸ See Arpaly & Schroeder (2014), especially their Chapter 6, for fuller explanations of the complexities involved in reward learning and desiring. Schroeder (2004) is also referenced there for even more detailed, and predominately neuroscientific, explanations and arguments of and for the reward theory of learning and the reward theory of desire that rests on it.

²⁴⁹ As far as I can tell, nothing about the reward theory of learning or desire requires solving the problem of mental causation, or even the problem of what the relationship is between mental and brain states. If epiphenomenalism is true, for instance, then we can talk about the brain or neural states that mental states supervene on causing one another.

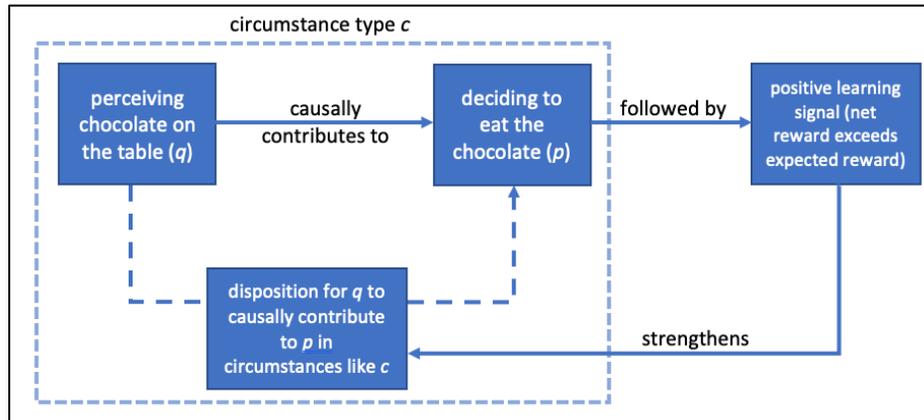


Figure 10: Illustration of (Part of) Reward-Based Learning

Importantly, learning signals are positive, neutral, or negative in virtue of their causal contributions to modifying the relevant disposition (or not), and not in virtue of *feeling* positive, negative, or neutral. That is, your positive learning signal that follows the decision to eat the chocolate is not the same thing as the pleasurable feeling you (might) get while eating it. This is why it is possible for the chocolate to taste delicious despite getting a negative learning signal – for example, because you also feel overwhelmingly guilty for breaking your diet by eating the chocolate. Now, the next question is: What determines whether a positive, neutral, or negative learning signal is released?

The simple answer is that representations of the world do. Some representations of the world (or what they signify) increase the chance of a *positive* learning signal, some increase the chance of a *neutral* learning signal, and some increase the chance of a *negative* learning signal.²⁵⁰ The learning signal is the outcome of a reward calculation, and so increasing the chance of some particular outcome means that there is some influence on the calculation process. You find yourself in a set of circumstances for which you have some expected overall reward value, and this *expected reward value* is then figured into the reward calculation and subtracted from the *actual net reward value*.²⁵¹ So, what determines the outcome of this calculation is a combination of how good (or bad) you *predict* the world to be, and

²⁵⁰ Again, the reader is referred to Arpaly & Schroeder (2014) and Schroeder (2004) for the details.

²⁵¹ The reward predictions and calculations are done unconsciously. Arpaly & Schroeder (2014, Sect. 6.1) explain this, as well as the fact that such calculations are as easy as they are common for our brains to perform on their own. They again point to Schroeder (2004) for readers inclined to want to know the underlying neuroscience.

how good (or bad) you *now take* the world to be. These both involve representing the world with some associated reward value (expected vs. actual). Thus, representations that increase the chance of a *positive* learning signal constitute *positive contributions* to the reward calculation; those that increase the chance of a neutral learning signal constitute *neutral contributions*; and those that increase the chance of a negative learning signal constitute *negative contributions*.²⁵²

The obvious next question is: What makes a representation of the world contribute in the way it does to the release of a learning signal? The answer is two-fold. First, sometimes this is innate. That is, some of our representations of the world will contribute positively, neutrally, or negatively by “nature’s design,” since we come into the world ready to represent “sweet taste experiences, full stomachs, [and] dry bottoms” as rewards.²⁵³ That is, such representations innately contribute positively to overall reward calculations. Some representations indicating danger, such as those of spiders, snakes, and large sharp-toothed cats, will innately contribute negatively to the reward calculation.

Second, experience can determine how a representation contributes to reward calculations. This is essentially done via association.²⁵⁴ Were you to be punched in the face each time you ate a piece of sweet-tasting food, you (or rather, your reward system) would quickly learn to treat representations of such food (and such tastes) as *punishments* – as negative contributions to overall reward. Conversely, if you have owned snakes since you were young, handled them regularly, and never really had any trouble with them, you may very well treat representations of snakes as rewards – as positive contributions to overall reward. Thus, whether innately or as a result of experience, it will simply be true for some individual that certain of their mental representations contribute positively, others

²⁵² For positive and negative learning signals, there is an additional detail. When an expected negative contribution (called a ‘punishment’) is *absent* (or removed), this can contribute positively to the overall calculation by *decreasing* the chance of a negative learning signal. Since they effectively increase the chance of a positive learning signal (a reward) in virtue of the absence of some punishment (a negative), these are called ‘negative rewards’. Conversely, *negative punishments* involve the absence of an expected positive contribution; being punishments, these increase the chance of a negative learning signal.

²⁵³ Arpaly & Schroeder (2014, p. 132).

²⁵⁴ Once again, Arpaly & Schroeder (2014) and Schroeder (2004) should be consulted for the fuller, detailed story.

neutrally, and others negatively to the overall calculation of reward value. Figure 11 below illustrates this idea of contributing to a reward calculation that then has some type of learning signal as its output.

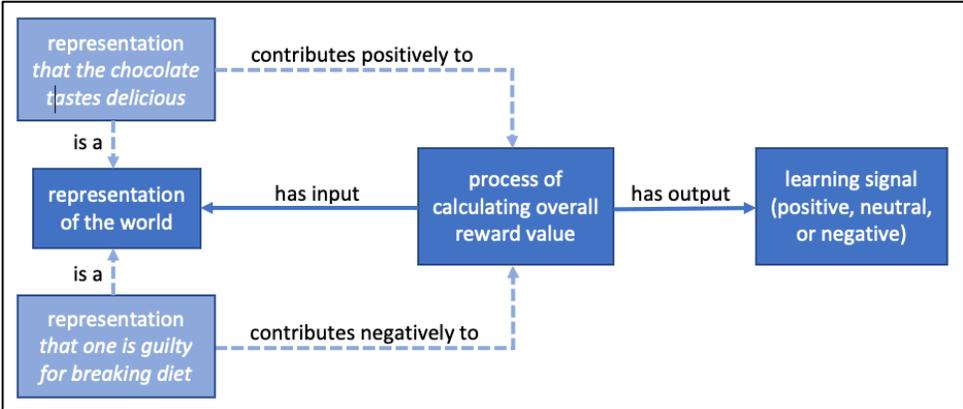


Figure 11: Illustration of (Part of) the Reward Calculation Process

Putting all of this together, for your reward system to treat the representation *that p* as a reward is for that representation to contribute positively to the overall reward value calculation. It is for that representation to figure into the calculation such that *the likelihood* that a positive learning signal is produced is increased. The contribution is still positive even if the overall learning signal ends up being negative (perhaps due to having more negative contributions), since it is only about that representation’s *contribution* being positive, neutral, or negative. The converse can be said about mental representations of *p* being punishments in virtue of their negative contribution to the overall reward value (and learning signal).²⁵⁵ Treating the mental representation *that p* as a punishment can be thought of as a desire *that not-p*. Following Arpaly and Schroeder, we can call this an ‘aversion’.

Desires, then, are *dispositions for mental representations of certain states of affairs (or what they signify) to be treated (by the reward and motivation system) as rewards*. Aversions are dispositions for such mental representations to be treated as punishments. So, to desire *that p* is to be disposed to have mental

²⁵⁵ The perception of the world after an act will be complex. We do not *merely* perceive the chocolate as tasting delicious, and so we do not *merely* represent some simple state of affairs. We may represent ourselves as having cheated on our diet or as feeling guilty, we may represent our action as right or wrong, and so on. It is the many representations in our post-act perception(s) that figure into the calculation of the overall reward value (and thus what kind of learning signal will be produced). I tried to capture this idea in Figure 11 by including two differently contributing example representations.

representations *that p* treated as rewards (by your motivation and reward system). *Desirings* are the realizations of these dispositions – such representations actually being treated as rewards by occurring and then contributing positively to the overall calculation of reward value. To desire eating chocolate is to be disposed such that mental representations of eating chocolate constitute positive contributions to overall reward calculations.

But learning signals follow causal sequences of one mental state bringing about (or causally contributing to bringing about) a second, and those learning signals either strengthen, do nothing to, or weaken the disposition of the first to bring about the second (in similar circumstances). Thus, desiring *that p* entails not only that mental representations *that p* increase the likelihood of a positive learning signal, but also that such representations increase the likelihood that the relevant disposition is strengthened.

3.4.1.1.2 Unconscious Desires, Habits, and Appetites

Two caveats. First, desires *do not* require conscious access while they are present or effective in action. When thinking of desires, it is common (perhaps because it is easy) to think of cases where some person has some desire *that p*, but where the desire is in full view of the agent. For instance, Shannon might desire *that she eat the chocolate bar on the table*, but where Shannon is also consciously aware of this desire, considers it, weighs it against other desires (such as maintaining her diet), and perhaps uses it (in quite a careful and reflective way) in her deliberation and subsequent decision about whether to eat the chocolate bar. This is, of course, an entirely plausible (even common) example of how some cases of desiring can unfold. Still, the familiarity or commonality of this type of case of desiring should not confuse us into thinking that, for any desire, it will be in full view of the agent in this way. I might desire to take a drink, or stand, or move the cursor on my laptop, or any number of

other things. Nonetheless, in desiring such things or even in acting on the basis of such desires, I am sometimes not at all consciously aware of the desiring or its influence on my subsequent action.²⁵⁶

Second, desires are *neither* mere tendencies to action (habits) *nor* mere tendencies to pleasure (appetites). Regarding the former, there are a number of differences between habits and desires: they are intuitively different; excusing practices distinguish them; habits do not engage our emotions, attention, or thoughts like desires do; and triggering conditions for habits are much more specific than for desires.²⁵⁷ This difference also still allows for desires to be intimately related to tendencies to act, such as by partly constituting or entailing them. Moreover, even if we supposed that some behavior is an action only in virtue of its relation to a reason, which plausibly involves some desire, this need not mean that something is a desire only in virtue of its relation to an action.²⁵⁸ For these reasons, desires and mere tendencies to act should be seen as distinct.

Regarding appetites, while pleasure (or displeasure) is a typical effect of getting (or not getting) what one desires, this tendency is distinct from what a desire is. One's desires might explain or typically cause this tendency, but there is evidence that they come apart. Patients on high doses of chlorpromazine, for instance, "are moved by their appetitive intrinsic desires, but no longer feel them... [so desires] are not essentially states that make people feel excited about their contents, or find their contents appealing."²⁵⁹ What is more, states of pleasure (or displeasure) *represent* desires, such as their satisfaction (or frustration). Hence, they could not be even partially constitutive of them, just as a photo of me cannot be partially constitutive of me.²⁶⁰

²⁵⁶ Note that this does not entail that the person cannot be made retrospectively aware of their desiring. That is, it might still be possible in these cases to point out to the person after the fact that they had such-and-such a desire, and for them to only then recognize on reflection that they had an experience or feeling of (or related to) that desire.

²⁵⁷ See Arpaly & Schroeder (2014, p. 112).

²⁵⁸ See Arpaly & Schroeder (2014, pp. 115–116).

²⁵⁹ Arpaly & Schroeder (2014, p. 118).

²⁶⁰ See Arpaly & Schroeder (2014, pp. 19, 21, 24–25).

3.4.1.2 The Role of Desires in Addiction

With an understanding of what desires are, let us turn to their role in addiction. I will focus on two important ways in which desires fit into the dispositionalist account of addiction.

First, desires – the disposition for mental representations *that p* to be treated as rewards – will serve to partially constitute the addiction disposition.²⁶¹ An addict will always have some desire to engage in certain types of behaviors, where the *certain type* will correspond to the type of addiction. For instance, an individual addicted to alcohol will have some alcohol-related desires, such as engaging in behaviors like seeking, obtaining, or consuming alcohol. Of course, other dispositions will also partially constitute the addiction disposition, such as the disposition to engage in exaggerated hyperbolic delay discounting.²⁶² The main idea is that, like my tennis serving ability, addiction is a complex disposition which will always rest on some type of arrangement of other dispositions.

My claim here is that desires will always be a part of that arrangement. No one is addicted *simpliciter*. Addicts are addicted *to* something, and my view is that this is properly understood as a type of behavior.²⁶³ Hence, for any addict, part of the arrangement of dispositions that constitute their addiction will be desires to engage in that type of behavior – the *dispositions* for mental representations

²⁶¹ As noted above, dispositions can occur hierarchically in the sense that some complex dispositions may depend on particular arrangements of other (sometimes lower level) dispositions. My tennis serving ability and most other skills are examples. I will speak of ‘constitution’, but I do not foresee it mattering whether this dependence relation is one of constitution, or grounding, or whatever else. Perhaps the sciences will eventually help determine this as they develop.

²⁶² Everyone discounts temporally distant rewards to some degree, and we do so hyperbolically, meaning that the rate at which we discount delayed rewards will change as temporal distance from the present increases. This is in contrast to exponential discounting, wherein one’s delay discount rate would remain constant over time. The upshot is that, in hyperbolic discounting, one’s preferences for inconsistent options can switch as one of those options (but not the other) becomes imminent. This is why people sometimes choose what they judge to be smaller, sooner rewards over what they judge to be larger, later rewards. Addicts have steeper discounting curves than non-addicts, meaning that perceived rewards lose their subjective value more quickly for addicts as they become temporally distant. This has the effect of speeding up the preference shift so that, essentially, a larger, later reward does not have to be as far into the future for an addict to switch preferences to a smaller, sooner reward. See Ainslie (2017, 2019) for discussion of delay discounting in addiction. See Appendix A for a bit more detail on delay discounting.

²⁶³ Possibly *types* of behavior, but I think it is more likely that an addict will ultimately be addicted to some type of behavior, such as consuming alcohol, and this addiction will manifest in all sorts of related types of behavior, such as seeking and obtaining alcohol. In short, when one is addicted to something, they are addicted to doing certain things involving or related to that something. Worries about babies being born addicted or the irreversibly comatose being addicted are not a problem when we realize that addiction is not the same thing as *dependence*. I discuss this difference more in Chapter 4.

of that type of behavior (and intimately related phenomena) to contribute positively to calculations of overall reward value.

Second, *desirings* – the *process* of such representations contributing positively to calculations of reward value – will be involved in important processes related to addiction, including its realization. But mental representations *that p* do not get treated as rewards in isolation. This is only one part of the ongoing feedback loop that is the operation of our reward and motivation system. We saw that an overall positive learning signal (your reward system saying, “Great, that was even better than expected! Do that again in the future!”) strengthens your disposition to behave similarly in similar circumstances. This involves positive contributions to the calculation of that learning signal (*desirings*). In addition, though, those positive contributions are also remembered as such. Thus, undergoing a process of desiring *that p* will, all else being equal, result in a greater chance of desiring *that p* later. In other words, it results in a greater chance of future mental representations *that p* contributing positively to the calculation of the overall reward value (and learning signal).

Moreover, our felt urges for one option or another experienced in the midst of decision-making (or even in less reflective moments of simple want-based actions) are driven by our *desirings*.²⁶⁴ The more that our mental representations *that p* are treated as rewards, and the more *strongly* they are treated as such, the more motivational appeal those mental representations will have. In this way, undergoing processes of desiring *that p* is important to addiction because this process helps to create the *cravings* experienced in addiction. In the context of reward learning, *cravings* constitute predictions of reward value – the more predicted reward, the stronger the craving. Recall that reward learning is largely experienced based. So, the more our mental representations *that p* contribute positively to calculations of overall reward value, the more your reward system will treat mental representations *that p* as something worth pursuing. This manifests itself in experience as craving, and it is exacerbated

²⁶⁴ See Arpaly & Schroeder (2014, pp. 96–98).

when mental representations *that p* have not merely contributed positively to the calculation of overall reward value but have also actually helped to produce an overall positive learning signal.²⁶⁵

Other processes that are influenced by desirings or have desirings as process parts include the *development* and the *strengthening* of an addiction. Strong cravings and strengthened dispositions to act on desires help to create and maintain an addiction. This is because they can influence and partly constitute the impaired control over one's desires that realizes an addiction. In a way, then, desirings reinforce the addiction disposition by reinforcing the impairment in one's control over their desires.

These are some of the roles that desires and desirings play in addiction. Desires partially constitute the addiction disposition. Desirings take part in the realization and reinforcing of an addiction. The latter point, having to do with realizations of addiction, brings us to impaired control.

3.4.2 Impaired Control as a Component of Addiction

3.4.2.1 Initial Motivation and the Misleading Disease vs. Choice Debate

At first glance, one might think that requiring impaired control as an essential component of addiction is much more controversial than requiring desires. One reason to think this is that the *addiction-as-disease vs. addiction-as-choice* debate appears to be raging on. In some sense, this debate is raging on. I think this is ultimately misleading, however, and so we should start by getting clear on this point. Doing so will serve as an initial motivation for thinking impaired control is central to addiction.

There are a minority of views defending something akin to the *moral model of addiction*.²⁶⁶ According to this kind of view, addicts have full control over their choices – most importantly, including *after* they are fully in the throes of an addiction – and the only thing they can really be said to suffer from is a defect in character. Addicts, they claim, make full, free, and responsible choices in

²⁶⁵ Recall that there can be positive contributions to the calculation of overall reward value even when the output of that calculation is a negative learning signal.

²⁶⁶ For instance, see Peele (1987), Schaler (2000), and Dalrymple (2008).

just the same way that you and I make choices about whether to order dessert, skip the gym, put off house chores, and so on. The only difference is that their choices are morally corrupt.

This is *not* the debate over whether addiction is a disease or a choice. To be sure, such views greatly influenced the initial rise and motivation behind the disease model sometime around the turn of the 19th century.²⁶⁷ The disease model is often reactionary to the ideas of the moral model, emphasizing that the disease model is the best way to combat stigmatism, blame, and other negative moral attitudes that are unduly directed at addicts because of the moral model's influence. Nonetheless, the moral model is now virtually unseen in mainstream debates over addiction's status as a disease and the role of control and choice in addicted behavior.

Instead, disease model proponents face off against proponents of what I refer to as 'choice models', where the latter defend a non-moralized view according to which addicts *retain some choice and control over their behavior*.²⁶⁸ Choice model proponents are often engaged in a kind of reactionary response of their own, seeing disease model proponents as overcorrecting from the moral model. As we might expect, this double overcorrection is partly responsible for the fishtail spin the literature is currently in. Choice model proponents tend to (understandably) see disease model proponents as arguing that addicts are *irresistibly compelled* such that their addiction undermines their agency and choice. Hence, the 'choice' in 'choice models of addiction' represents a negation of disease model's overcorrection. It means that addicts are *not* irresistibly compelled non-agents in virtue of their addiction. This, of course, leaves plenty of room for impaired control in the choice models – room that many choice model proponents are keen to fill. Thus, while they emphasize that addicts are not robotic non-agents, choice model proponents still acknowledge that something, to some degree, is amiss with an addict's ability to reign in their behavior through normal mechanisms of choice and control.

²⁶⁷ See White (2000).

²⁶⁸ Many examples exist in the literature (Heyman, 2009; Levy, 2006; M. Lewis, 2015; Pickard, 2019; Wakefield, 2017a; West & Brown, 2013).

As we can see, this debate is often confusing and misleading. With the exception of some outlier views, basically everyone seems to agree that there is some sort of impairment in control. The question is just how much. Unfortunately, this point is often overshadowed by choice model proponents' focus on undermining the disease model image of addicts as helpless non-agents. I will evaluate these views in more detail later. For our present purposes, just note that despite appearances, requiring some degree of impaired control as an essential component of addiction is neither inconsistent with any mainstream views in the literature nor a commitment to one side or the other in the *disease vs. choice debate*. Impaired control is still a unifying feature of most accounts across the board.

3.4.2.2 Further Reason to Require Impaired Control

Let us turn now to some further reasons for thinking that control is impaired in addiction. Consider first ordinary observation. While anecdotal, it is still worth noting that most people have seen (or heard legitimate stories of) addicts struggling to control their choices and behaviors (even if they *sometimes* can). Even choice model proponents admit that the disease model provides a simple explanation for the *puzzle of addiction*: addicts continue to use despite harm and other negative consequences because, to a significant degree, they cannot help it. Pulling back the reins on addicts not being able to control themselves does not fully undermine the utility of this explanation. We see and hear of addicts struggling in many ways related to control, such as maintaining their health, jobs, homes, families, and so on. A simple explanation is that their addiction has impaired their control.

Second, there is a wealth of empirical evidence that addicts suffer impaired control as compared to non-addicts. For instance, addicts consistently show impairments to areas in the prefrontal cortex associated with self-control.²⁶⁹ Relatedly, addicts also exhibit an impaired ability to inhibit impulsive behavior (relative to non-addict controls).²⁷⁰ It is also well-known that addicts exhibit

²⁶⁹ For instance, see Volkow & Fowler (2000) and Goldstein & Volkow (2002).

²⁷⁰ See de Wit (2009) for a review.

exaggerated delay discounting.²⁷¹ This means that, in general, when choosing between competing options, addicts find it harder to maintain a *global* perspective wherein their long-term, time-independent preferences are kept in view. Combined with the previous evidence, this suggests that addicts are more now-oriented due to an impaired ability to exhibit more reflective control over their impulses. This is so even when the cost is losing out on something that they judge to be more valuable than what their impulses are pushing them towards.

Third, as was alluded to above, arguments proposed against disease models (including those from both choice and moral models) appeal to evidence that is consistent with control being impaired in addiction. For instance, it is often pointed out that many addicts eventually recover without treatment and that they are responsive to incentives and other features of their environment.²⁷² Recall that arguments like this are typically put forth as a way to combat disease model proponents' (purported) claims that addicts are irresistibly compelled non-agents. However, these arguments could only show that addicts do not *fully* lack control or that control is not impaired at *every* moment. I argue for neither claim, nor does defending an impaired control condition for addiction require these claims.

Fourth, consider a case in which an individual has full – again, *full* – control over their choices, but at each moment of choice over whether to engage in some dug-consuming behavior, they decide to do so. If each choice really were fully up to the person, such that there is no deficiency in their control whatsoever, then it is hard to see why this person should count as an addict. Stop for a moment, close your hand into a fist, and then, when you decide, open it again. Your decision to open your hand seems like a paradigm case of full control. An individual with control like this – full, unadulterated control – over their (purportedly addicted) behaviors is, intuitively, no addict at all.

²⁷¹ See Ainslie (2019) and Heyman (2013) for discussion. See Appendix A for detail on hyperbolic delay discounting.

²⁷² See, for instance, Heyman (2009, 2013), Pickard (2019).

However, an objector might think that addiction is only about desires, and still want to call such a person an addict if it remains true that they regularly experience strong urges to engage in a certain type of behavior. The ability to control those urges, the objector argues, is neither here nor there since addiction is just a matter of strong, persistent urges (perhaps in a disordered way). There are a few things to say. First, this response risks conflating intrusive thoughts with addiction. To be sure, addiction involves persistent desires (and so urges) in some sense because addicts will be strongly disposed to experience them in predictable circumstances. But the mere presence of (or disposition towards) persistent desires and urges is common to such conditions as obsessive compulsive disorder, attention deficit hyperactive disorder, depression, anxiety, body dysmorphia, psychosis, post-traumatic stress disorder, and others.²⁷³ Hence, persistent (even disordered) desires cannot be sufficient for addiction without entailing that these disorders involve an addiction. Second, intrusive thoughts often involve desires or urges, which side-steps any claims to a disanalogy here. Third, the objector would owe us a story about how disordered desires *that are addiction* (on their view) differ from disordered desires *that are parts or symptoms of the above disorders* (manifesting as intrusive thoughts in those disorders). Fourth, persistent (or disordered) desires are likely going to be associated with impaired control in any case, as they are in the above disorders.²⁷⁴ Consider that we can typically indirectly control the occurrence of our desires by working toward minimizing or eliminating them. So, if one cannot do this, it seems that this already involves some sense of impairment of control over their desires, which might be contributing to the fact that they are disordered in the first place.

Having laid a solid foundation for including some degree of impaired control in addiction, let us turn to an explanation of this component as *the systematic loss of control* on the dispositionalist account of addiction I defend here.

²⁷³ This is clearly seen in the overlap of the criteria for numerous mental and behavioral disorders in the DSM-V (American Psychiatric Association, 2013) and the ICD-11 (World Health Organization, 2020). See also Winston & Seif (2017).

²⁷⁴ Again, see Winston & Seif (2017).

3.4.3 Addiction as a Systematic Loss of Control

Addiction is realized in failures to control one's desires. Control comes in degrees, and many non-addicts often fail to control their desires some of the time (just as non-fragile objects often break). So, the natural next question is: In what sense must control over one's desires be impaired to count as an addiction? There will be two parts to the answer to this question. The first concerns *degrees* of control and revolves around the notion of *systematicity* introduced above. This is a core piece of the account. The second concerns *types* of control and it attempts to explain how the dispositionalist framework offers a way to capture impaired control in addiction that is neutral between competing *accounts* of control and consistent with different *types* of control.

I start by considering degrees of control and systematicity. First, recall from above that a disposition is systematic when it is sufficiently strong – that is, sufficiently likely to be realized when triggered – in a sufficient number of those triggering conditions which members of the bearer's reference class have a sufficiently high likelihood of being in, given the laws of nature and history of the world. We now need only to apply these features to addiction, and then add in some clarification for good measure. I will start with the latter, briefly saying more about what it means for an addiction to be realized in failures to control one's desires. This will help us to understand how to make sense of an addiction being *sufficiently strong* in sufficiently many of the *relevant triggering conditions*.

3.4.3.1 How We Can Fail to Control Desires

Desires are dispositions for our mental representations of certain states of affairs to be treated as rewards (by our reward and motivation system). In addition, a desire *that p* influences our experiences regarding *p*, most saliently by modifying our dispositions to feel pleased (or not) at *p* being actualized (or not), and to feel more or less moved to pursue actualizing *p*. More than this, desires can directly influence our thoughts and actions by modifying our dispositions to think or act in particular

ways that promote actualizing p . In other words, desires regarding p are intimately connected to felt urges regarding p as well as tendencies to thought and action regarding p . How, then, do we fail to control our desires?

The paradigm case is failing to resist a present motivation or felt urge to do something, which exists in virtue of some desire for doing that thing. Suppose Jim strongly desires to use heroin and is currently presented with an opportunity to do so. Given his desire, Jim feels a strong urge to go ahead and use heroin. However, given some of Jim's other desires, he also feels motivated, albeit less so, to resist using heroin in this instance. Perhaps Jim knows using heroin will devastate his mother and risk him losing his basketball scholarship. Thus, Jim is somewhat torn, given that he does desire to use heroin (there is nothing new or special about a person's being torn, nor about there being reasons – even good ones – to sometimes use drugs). Ultimately, though, in this instance Jim fails to resist his urge to use heroin. This might take a number of different forms, such as the following:

- (i) Jim has trouble holding in his mind the thoughts that counteract his motivation to use heroin (perhaps he is easily distracted by the heroin urge and its corresponding cues);
- (ii) Jim finds it easy to think of countervailing reasons for those thoughts, even if he can hold them in his mind (such as doubts that his mother will ever find out or the thought that he never really wanted to be a professional basketball player anyhow);
- (iii) Jim underestimates the efficacy of his heroin urge and actually attempts to *simply resist it* – through something like a simple, non-deliberative mental effort – yet, due to his capacities, he fails in this effort;
- (iv) Jim remains torn despite some deliberation, only to have the motivational balance tipped by some moderate urging from his friends (“Come on, Jim! This will be the *last* time!”);
- (v) Jim simply fails to deliberate at all and gives in almost immediately to his heroin urge, despite prior resolutions to resist doing just that (he at least wanted to deliberate).²⁷⁵

We have various mental and behavioral capabilities that are involved in motivation, deliberation, decision-making, self-reflection, modification of our desires and dispositions, and so

²⁷⁵ Again, remember that in each of these cases Jim *attempts but fails* to control his desires, and so the failure here is going to be some positive – yet uncontrolled – event, such as a behavior or choice. It is easy to describe events as either essentially an absence (when you do x , you fail to do *not-x*) or essentially positive (when you fail to do x , you do *not-x*).

forth. I will call these ‘cognitive resources’.²⁷⁶ Desires influence our felt urges and motivations, and so one way to control our desires is to utilize these resources in trying to resist those urges or motivations when they arise.²⁷⁷ Hence, one way to fail to control our desires is to fail to resist a felt urge or motivation due to a failure to implement (or to *successfully* implement) some such capacity, just as Jim failed to do in the ways listed above.

Now, people are generally not very good at actively resisting felt urges that they want to resist.²⁷⁸ Hence, another (probably better) way to control our desires is more indirect. Instead of waiting for a desire to become active in the moment – for instance, by producing a felt urge when triggered – we might take antecedent steps to modify our desires. Modifying our desires can range from eliminating (some of) them entirely to reducing their influence. One way to do this is to utilize the aforementioned cognitive resources, but to doing so indirectly involves utilizing them in contexts which precede the triggering of our desires into felt urges. For instance, Jim could try to become more aware of (even vigilant about) which cues and circumstances are likely to lead to opportunities (and felt urges) to use heroin. In general, resisting an urge to go to a friend’s house where heroin might be present will be much easier than resisting the urge to use heroin once it is offered to you at that friend’s house. Jim could also try to strengthen his anti-heroin-use desires, for instance by imagining his mother finding out or conditioning himself to picture her crying to him to quit when he thinks of heroin or his heroin cues (like being invited to his friend’s house where there may be heroin).

So, we can utilize these resources antecedently in various ways. This will involve implementing them in non-triggering contexts, but this implementation might take different forms. We can add to

²⁷⁶ Though I include behavioral capabilities, the idea is that we have certain mental capabilities that are involved in influencing both how we think and how we subsequently act. Hence, they are *cognitive* resources in the sense that their influence (whether on thoughts or actions) initially stems from our ability to think in certain ways.

²⁷⁷ Unfortunately, not wanting to do *p* for whatever reason (and so having some desire to *not-p*) is insufficient for lacking desires to *p*. We are often torn about what we want to do or, at least, often have inconsistent desires. See Arpaly & Schroeder (2014), especially Chapter 2 on deliberation.

²⁷⁸ See Baumeister (2018).

our cognitive resources by practicing deliberation, self-reflection, attentional control, and the like, such that we gain cognitive capabilities we did not previously have. We can also strengthen the resources we already have such that their later use is more effective or consistent – perhaps by making it possible (where it otherwise was not) to anticipate and then preempt the relevant urge. We can also build up certain urge-countering tendencies that will be triggered in the relevant circumstances.

Hence, controlling our desires is not simply about attempting to resist a presently felt urge that our desires give rise to. It will sometimes be about this. But it will sometimes be about making more indirect efforts to control our desires, such as taking steps to eliminate or mitigate them, to enhance others countervailing desires, to build up our cognitive resources, and so on.

3.4.3.2 Systematicity and Societally Relevant Circumstances

Of course, one failed attempt to control our desires, whether direct or indirect, is insufficient for addiction. This means that addiction is distinct from one-off cases of weakness of will. On my account, the difference can be explained by appealing to systematicity.

Let us start by noticing something important about those who argue that there is no difference in kind between addiction and weakness of will. Such views claim that there is no categorical difference in terms of the failed exercise of control between an addict and someone performing an *akratic* or weak-willed act.²⁷⁹ The point these authors emphasize is that nothing special happens (in terms of failed control) when an addict fails to resist an addiction-produced urge. It is just like when a non-addict tries and fails to resist an urge in a single bout of weakness of will, as when I reluctantly give in to my midnight snack urge, eat the cheesecake, and quickly regret it. We have all likely suffered from weakness of will on some occasion, and so, on these authors' views, we should all know what it is like for an addict to succumb to their addicted cravings. At least, this seems to be the idea.

²⁷⁹ For instance, see Heather (2017a, 2020) and Foddy and Savulescu (2010a).

Notwithstanding this purported similarity between addicted and weak-willed behavior, there is still an important difference that these accounts allow: frequency. According to these accounts, while addicts succumb to their desires just as *akratics* do, addicts do so with much higher regularity. On my account, this is best understood as a difference in capacity, and it gives the game away to systematicity (and, at the very least, dispositionalism). Addicts are *systematically disposed* to fail to control their desires and urges, and this explains the difference in regularity between an addict and an ordinary weak-willed non-addict. We now just need to see what this difference in regularity comes to.²⁸⁰

Our discussion of systematicity above is helpful here. First, the disposition to fail to control one's desires needs to be sufficiently strong. But the *no difference* theorists would hold that an addict's disposition is not necessarily stronger than a non-addict's disposition to fail to control their desires (at least not in virtue of having an addiction). However, the evidence discussed above for the inclusion of impaired control in addiction also supports the rejection of this claim that there is no difference between addicts and *akratics*. Regarding impaired control, there are statistically significant differences between addicts and non-addicts (who very likely perform some *akratic* actions). Addicts exhibit weaker abilities to inhibit impulses and delay immediate gratification, for instance. These data suggest that, *pace* the no difference theorists, there *is* a difference in strength of the disposition to fail to control one's desires. That is, addicts' dispositions are more prone to being realized when triggered than are mere *akratics*, and this can be explained by the evidence that there are statistically significant differences in their cognitive arsenals (both in terms of quality and available inventory).

Even still, for the disposition to fail to control one's desires to be systematic, it is not enough that it is sufficiently likely to be realized when triggered (recall strong yet non-systematic dispositions

²⁸⁰ Note that it is consistent with my view that there can be brief addictions. The idea is basically that the *akratic* and the brief addict will differ in their dispositions. A brief addict will be *systematically* disposed to fail to control their desires, and the *akratic* won't. This is roughly the difference between fragile vases and bowling balls that break when dropped from very tall buildings. The brief addict would simply acquire the systematic disposition and then lose it very quickly.

like our disposition to suffocate in the stratosphere). We need to contextualize that strength to certain types of triggering conditions. As the reader will recall, systematicity is partly about which triggering are relevant towards determining whether a disposition to x is indeed *systematic*. Systematicity helped explain why dropping bowling balls or glass vases from very tall buildings was not the right kind of test for fragility. Of course, performing this test would trigger the vase's fragility, producing its realization.²⁸¹ But doing so would also trigger a disposition of the bowling ball to break, though not the systematic disposition to break that is *fragility*. Thus, it would be an inconclusive exercise, and part of the reason is that this type of circumstance is not relevant in the right way. That the vase can break when dropped from the Empire State Building is a way the vase is, to be sure. Similarly for the bowling ball. Still, the vase and the bowling ball being disposed to break in *that* sense is not what we mean when we call the vase 'fragile' and the bowling ball 'not fragile'. To get at what we mean, we need to restrict the circumstances to those types that are relevant – this is what systematicity is doing.

So, which are the right types of circumstance to test for an addiction? The general answer will be familiar: *those for which members of the addict's reference class have a sufficiently high likelihood of being in, given the laws of nature and history of the world*. Following David Limbaugh's discussion of systematicity in the context of delineating dysfunction and disorder, I will call these circumstances 'societally relevant'.²⁸² We need a bit more, though. Just as fragility can mostly be restricted to certain types of striking, we can mostly restrict the addiction-relevant circumstances to what I will call 'choice situations'. By this I mean that societally relevant circumstances need to involve the individual being presented with an opportunity to make an addiction-relevant choice. This need not be the choice to use or otherwise engage in the addicted behavior. It might simply be the choice to put oneself in a position to make

²⁸¹ As a side note, I am still unsure whether I think that this should actually count as a *triggering condition* in any meaningful sense. If we exploded the vase with a nuclear bomb (and everything else within a 10-mile radius), it is unclear to me whether this would properly be called a 'realization' of the vase's fragility. Realizations *reveal dispositions for what they are*. I am more inclined to say the nuclear blast – given its effects on everything else – would not be doing this. The same might be said for dropping it from a very tall building, given what this also does to bowling balls, pianos, and the like.

²⁸² See Limbaugh (2019, p. 12).

that choice. It may also be a choice about whether to sign up for the local AA meetings, to call and make an appointment with a therapist, to throw away one's cigarettes, to confide in a friend so as to establish something of a support system, and many other types of circumstance that are connected to an eventual choice to engage in the addicted behavior.

These circumstances and many others like them are societally relevant in the sense described above (members of the reference class are sufficiently likely to be in them). Given the above account of systematicity, the questions relevant to determining the presence of an addiction concern:

- (i) whether an individual is even disposed to fail to control their desires to engage in certain types of behavior; and,
- (ii) if so, whether their disposition is sufficiently strong in a sufficient portion of these societally relevant choice situations.

If both are true, then the disposition is an addiction (since it would be the *systematic* disposition to fail to control one's desires). If not, it is not. We can think of the addiction test as one of considering someone with this disposition being embedded into each possible societally relevant choice situation. The test is meant to determine whether the disposition is systematic. Suppose Al would fail to control his desires 85% of the time in 85% of these possible scenarios. This would mean that were Al to try to control his desires to, say, drink alcohol, he would almost always fail to do so in almost all of the relevant kinds of situations. This would be true *even if he never actually tries to control them* (perhaps he is a willing alcoholic). If Al were to lose his addiction, it would follow that these odds would change.

That is how systematicity is applied to addiction. It is the same basic framework we saw above with fragility, though the realizations and the triggering conditions changed. Fragility is the disposition to break when suitably struck. Addiction is the disposition to fail to control one's desires when presented with a (societally relevant) choice situation wherein one has the opportunity to indulge those desires and attempts not to. More specifically, it is being *strongly disposed* to fail to control one's desires

in such circumstances, and moreover being strongly disposed *in a sufficient portion* of such circumstances. Addicts have this disposition, with this sort of nature, while non-addicts do not.

3.4.3.3 Types of Control: A Theory-Neutral Framework

Now that we have filled out the content for the dispositional account of addiction defended here, I need to say something about accounts of control before summing up. Specifically, an objector may worry that I have not provided any specific theory of control that is to be used within my dispositional account of addiction. This has been intentional.

The framework just laid out is neutral between competing accounts of control. Moreover, it would not matter if these accounts were inconsistent with each other, but it is important to note that this is almost never the case. At least, this is almost never true in the sense that would disallow a theorist to utilize multiple such accounts within the dispositionalist framework. That it is neutral even with respect to inconsistent accounts of control makes the theory-neutral claim I am making even more plausible. Let me explain.

Theories of control are relevant to many areas in philosophy, but I will focus on free will and moral responsibility. Theorists in this domain disagree over the answer to many control-related questions. What does control come to? Would the truth of determinism – that the past and the laws of nature together entail only one possible future at each moment – undermine control, and thus free will? Does moral responsibility even require free will, and thus control, such that determinism might threaten free will and moral responsibility together? Unsurprisingly, there are a number of competing answers to the first question. The second question has only two possible answers (“yes” or “no”) but many different explanations for why each might be given. The same is true for the third question.

In such debates it can appear that competing answers must be entirely incompatible. My claim is that if we focus just on the accounts of control, we will see that these appearances are deceiving.

Consider how a compatibilist – someone who answers “no” to the second question – might answer the first question about what control comes to.²⁸³ They might contend that control amounts to *being appropriately sensitive to reasons, exhibiting sane choice patterns, hierarchical alignment between first- and second-order desires*, or some other determinism-friendly account about identifying the right sort of *causal sequence*.²⁸⁴ Incompatibilists, on the other hand, might contend that control amounts to *exercising an agent causal power, performing a self-forming action, being the non-causal subject of one’s action*, or some other determinism-unfriendly account of control.²⁸⁵ And since incompatibilists hold that determinism would undermine the ability to exercise these forms of control, one might think that there is some condition built into them – such as *being able to do otherwise* or *having multiple live possible futures* – that makes them inherently inconsistent with the compatibilist accounts of control. This is where we are deceived.

Compatibilist and incompatibilist accounts of *control* are not inherently inconsistent. If they were, then they could never consistently be part of a single theory. This is false. For instance, a compatibilist could hold that both reason-responsiveness and exercising agent-causal powers are *sufficient* for free will and moral responsibility. Alternatively, an incompatibilist could hold that these are both *necessary* for free will and moral responsibility. This would not be possible if these accounts of control were inherently inconsistent with one another. Most responsibility theorists hold that *some* degree of control over an action is required for an agent to be free or morally responsible for it. Typically, though, disputes over how much control is required come to *how* the agent controls their action. One does not often find theorists arguing over the need for 60 as opposed to 75 percent control. As the foregoing illustrates, the arguments are instead about which kinds of control are either necessary or sufficient (or not) for being free and responsible. It does not follow from this that one

²⁸³ This is a rough definition of ‘compatibilism’, but it will do for our purposes.

²⁸⁴ See Fischer (1982), Wolf (1990), and Frankfurt (1988), respectively, for such accounts. The phrase ‘causal sequence account’ comes from Sartorio (2016) and is meant to pick out certain compatibilist accounts of control. This is because these compatibilists focus on making control consistent with causal determinism, where there is only one possible causal sequence extending into the future, by claiming that sufficient control is about identifying the right kind of causal sequence.

²⁸⁵ See Lowe (2010), Kane (1998), and Ginet (1990), respectively, for such accounts.

or the other side must say that their competitor's account of control is not actually control at all, nor that does it follow that their respective accounts of control are incompatible. Even many skeptics and impossibilists about free will and moral responsibility hold that differing degrees of control are present in the world.²⁸⁶ The disputes, then, come to the sufficiency of one or the other accounts of control for establishing free will or responsibility.

This is important because it explains how a framework like mine, which leaves open what control comes to, could be filled in by a compatibilist account of control, an incompatibilist one, or both. For instance, failing to be appropriately responsive to reasons might be one way in which an addict can fail to control their desires. In fact, some addiction theorists have claimed just this.²⁸⁷ Alternatively, addicts may fail to control their desires due to a lack of or impairment in the ability to exercise their agent-causal power. Further still, as we just saw, there is nothing incoherent about holding both of these views. Hence, it is possible to utilize multiple accounts of control in filling in this further detail of the framework, and even those accounts which might appear at first to be incompatible.

Finally, let us come back to the stronger claim made above, that *even if* some of these accounts were inherently inconsistent, this would not matter to my claim that the dispositionalist framework is theory-neutral regarding accounts control. This would only mean that a theorist taking up my framework could not fill in the impaired control component using *multiple* such accounts. Suppose reasons-responsiveness and agent-causal powers were inconsistent accounts of control. (I cannot think of a reason why this would be true, which also speaks against this worry.) In that case, all that would follow is that addicts could not fail to control their desires in both of these ways (since both

²⁸⁶ A skeptic holds that, because of how our world is, free will and moral responsibility can never be instantiated in the actual world. An impossibilist holds that skepticism is true, and the skeptical view is true of every possible world. See Kelly & Kershnar (Forthcoming), Levy (2011a) and Pereboom (2016) for examples.

²⁸⁷ See Levy (2011b) and Uusitalo (2011).

accounts could not be true). Hence, if a theorist liked both of these accounts of control and the dispositionalist account of addiction defended here, then they would simply have to choose which one to use to explain addicts' failures to control their desires. They could not hold both as part of the same view, but each one by itself would still be compatible with the dispositionalist account of addiction. Hence, the latter is theory-neutral with respect to accounts of control.

As I see it, it is much more likely that almost every plausible account of control on offer can work well with the others. Addicts will fail to control their desires in many ways, fitting different accounts of control. That my framework can capture this makes it very useful indeed.

3.5 A Recap of Chapter 3's Arguments

I called this chapter 'stage-setting' at the outset. This is because the rest of the defense of premise (3) in the next chapter will rest on the work laid out above on unification, dispositions, and the account of addiction as the systematic disposition to fail to control one's desires. However, I still offered a series of arguments for the components of the stage I have been setting in this chapter.

First, I argued that ontology can provide only *methodological unification*, and not *substantive unification*. This had to do with its methodology being substance-independent, thanks to principles like *perspectivalism*, *adequatism*, and *AAA*. The promise that the dispositionalist account will supply this second type of unification will be kept in the following chapter.

Second, I argued for a particular understanding of dispositions, based in work by Neil Williams and Barry Smith, though without painting myself into an overly peculiar or idiosyncratic metaphysical corner.²⁸⁸ All one really needs, I argued, is that dispositions fit somewhere into their ontology. This is quite a minimal ask. I then provided definitions for 'disposition' and a handful of related dispositionalist terms, such as 'power', 'capability', 'function', and 'predisposition'. Finally, I argued

²⁸⁸ This is not at all to imply that Williams or Smith have done this – it is simply to make it clear that I have not done so.

that dispositions have at least five central features that are relevant to the account I defend (and probably to a dispositionalist account of anything, it seems). These were *triggering conditions*, *realizations*, *manifestations*, *material basis*, and *reliability*. The latter had three senses: *strength*, *opportunity*, and *systematicity*.

Third, I argued for the inclusion of three main components in the dispositionalist account of addiction: *desires*, *impaired control*, and *systematicity*. I offered an account of the nature of desires based in Nomy Arpaly and Timothy Schroeder's *reward theory of desire*, and I explained the role of desire in addiction. I then explained how the core component, *systematicity*, was applied to addiction. Lastly, I argued that the account is theory-neutral with respect to control. That is, the dispositionalist framework can be used by theorists with competing (even inconsistent) accounts of control.

With the dispositionalist meat now on its bones, let us turn to Chapter 4 where the framework is compared to others in the literature. There we will see how the account I have been defending fits with what we know empirically about addiction, how it relates to other accounts in the literature, ultimately providing unification to them, and how it would fit into a BFO-conformant ontology of addiction. Doing so will constitute the defense of premises (3b)-(3d), and thereby complete the defense of premise (3) of my overarching argument.

Chapter 4: The Power of the Dispositionalist Account of Addiction

4.1 Introduction and Chapter Road Map

Chapter 3 set the stage for displaying the power of the dispositionalist account of addiction. The task of the present chapter is to now display that power. To do so, I argue for the rest of the claims supporting premise (3) of my overarching argument. As a reminder, here is that argument again:

- (3a) Ontology cannot *fully* unify the literature (that is, there is a further, distinct sense of ‘unification’ that ontology alone cannot provide); [Chapter 3]
- (3b) The dispositionalist account of addiction best explains the key phenomena surrounding addiction;
- (3c) The dispositionalist account of addiction captures other competing theories and can explain their (often verbal) disagreements;
- (3d) The dispositionalist account works with the methods of ontology;
- (3e) If (3a)-(3d), then the dispositionalist account of addiction is true and provides still further unification to the literature;²⁸⁹
- (3) So, the dispositionalist account of addiction is true and provides still further unification to the literature. [(3a)-(3e)]

Chapter 3 defended premise (3a). This chapter defends premises (3b)-(3e).

I start with premise (3b) in Section 4.2, arguing that understanding addiction as a disposition best explains the key phenomena surrounding addiction. These phenomena include such things as: the so-called ‘puzzle of addiction’; the fact that addiction involves both impairment of control and an ability for choice (what I will call the ‘Janus-faced’ nature of addiction); the existence of a gray area of addiction (which can be understood in terms of ‘vagueness’ of the addict/non-addict boundary); the context-sensitive nature of addiction; the wide variety of types of addiction; and the polysemy in the

²⁸⁹ If (3d) were not true, then the dispositional account would not provide *further* unification to the literature since, *ex hypothesi*, the substantive unification it *could* provide would not be compatible with also providing methodological unification through ontology. Hence, for the dispositionalist account to add *further* unification to the literature, it must be compatible with implementing the methods of ontology. In one sense, this is easy. However, as I explain below, this is more about the dispositionalist account – and mine in particular – being compatible with BFO-conformant ontologies.

colloquial use of ‘addiction’ and its cognates. I will also argue that the dispositionalist account best explains four metaphysically possible cases of addiction.

In Section 4.3, I shift the focus to premise (3c). There I argue that the dispositionalist account can capture extant competing views, and moreover, that it can explain some of their disagreements with one another.²⁹⁰ Sometimes these disagreements end up being verbal, a point that the dispositionalist account can bring to the surface. I hope to make two things clear. First, the dispositionalist account narrows in on what many accounts are (usually implicitly) already getting at. Second, it is consistent with any plausible account in the literature. Not only is the dispositionalist account *consistent* with most extant accounts, but it can also help *explain* where and why they differ. This is an example of how the account is substantively unifying. This brings us back around to premise (3a), and into the next section.

Section 4.4 starts by making good on the promise from the preceding chapter to tie premise (3a), which was defended there, back into the argument for premise (3). I explain why, having defended premises (3b) and (3c), it should be easier to see why the dispositionalist account can be said to offer further unification *over and above* the implementation of ontology itself (which is just premise (3a)). Having brought us back to ontology, this section then explains and defends premise (3d), which states that the dispositionalist account works with the methods of ontology. If the conclusion of Chapter 2 is true, and of course I think it is, the substantive account of addiction I defend ought to be workable within that methodology. More than this, it ought to be *useful* within that methodology. The further unification the dispositional account provides goes some way towards such usefulness.

In Section 4.5, I bring the foregoing pieces of the argument together in a defense of the bridge premise (3e). Though this premise is not trivial, it should be a relatively smooth transition from the

²⁹⁰ See Section 3.2.1 in Chapter 3 if a reminder is needed for what I mean by ‘capture’ here. Roughly, other accounts can be incorporated into the dispositional framework, taking on what they get right and avoiding what they get wrong.

work done in the preceding sections to the truth of (3e). There are two brief parts to this. First, I explain why the truth of premises (3b) and (3c) support the truth of the dispositionalist account of addiction. Second, I explain why the truth of premises (3a) and (3d) support the claim that the dispositionalist account provides still further unification to the literature. In the end, we will see that premise (3) of my overarching argument is well-supported.

Section 4.6 is brief, summarizing the main lessons before transitioning into the final, comparative component of the project that is Chapter 5.

4.2 Capturing Key Phenomena with the Dispositionalist Account of Addiction

In this section, I argue that the dispositionalist account of addiction best explains several key phenomena surrounding addiction. The strategy here is as follows. I consider a series of facts about addiction that should be explained by any good account. For each, I argue that if we assume that the dispositionalist account is true, this either best explains that fact or, in some cases, is consistent with the best explanation. Some of these phenomena, such as the so-called ‘puzzle of addiction’, serve as the basis for the *divergence* between competing accounts seen in the literature. Therefore, that my account either best explains or is consistent with the best explanation of these facts is further evidence in favor of adopting it over extant accounts. Section 4.3 will add to this evidence by showing that these divergent views need not be thrown out entirely, since the dispositionalist account can explain their divergence in a way that ultimately unifies them.

The following subsections address these key phenomena in turn. First, a brief caveat is in order. The end of Section 4.2 discusses four metaphysical possibilities concerning addiction, such as the possibility that aliens could be addicted. Thus, an objector might resist my calling these ‘facts’ or my claim that they are worth being explained. First, that these are *metaphysical possibilities* does not exclude them from being facts – they are simply *modal* facts. The lack of concern in the literature with

such possibilities – especially, but not only, the clinical and experimental literature – only implies that many researchers simply do not care about these modal facts. However, this does not entail that they are not facts about addiction to be explained.

Second, we *should* explain such metaphysical possibilities if we are interested in *what addiction is*. By ‘explain’, all I mean is that one’s account should be consistent with these possibilities (that is, assuming the given account is true would explain how such cases are indeed possible). Moreover, we can think of capturing such possibilities as having additional marks added to the account’s ledger. Thus, if an account can capture these metaphysical possibilities concerning addiction without significant (or any) cost, then it seems to me that this speaks in favor of that account over others that fail to capture them. I aim to show that this is the position we are in with my account *vis-à-vis* many other extant accounts on offer. Let us now turn to the facts to be explained.

4.2.1 The Puzzle of Addiction

If there is any fact about addiction that everyone agrees must be explained by any good account, it is the *puzzle of addiction*. Thus, this is where we ought to start. Hanna Pickard provides a nice description of this puzzle:

Common sense suggests that if a person knows that an action of theirs will bring about negative consequences and they are able to avoid doing it, then they do. We act, so far as we can, in our own best interests and the interests of others we care for. This is a basic folk psychological rule of thumb for explaining and predicting human action, ubiquitous in our ordinary interaction with and understanding of each other. But this is what addicts seem not to do. Although addiction has severe negative consequences, addicts continue to use drugs. This is the puzzle of addiction: *why do addicts keep using drugs despite negative consequences?*²⁹¹

As noted, this puzzle is the basis for a major divergence between *brain disease models* and *choice models* of addiction. The former, as Pickard goes on to point out, provide a very simple explanation for the puzzle: addicts suffer from a compulsion-inducing brain disease undermining their control

²⁹¹ Pickard (2019, pp. 9–10).

over certain types of behavior, and so they cannot resist using even despite harmful and sometimes devastating consequences that they themselves want to avoid.²⁹² The choice models, on the other hand, deny compulsion and so must look elsewhere for an explanation. This can change the puzzle into, as Pickard puts it, “a puzzle of choice.”²⁹³ In other words, choice model proponents ask: Why do addicts *choose* to continue their addicted behavior despite negative consequences?

If we accept Pickard’s “rule of thumb” (from the passage above) about human behavior, this is certainly a harder puzzle to solve. Indeed, choice theorists can end up trying to *explain away* the puzzle of addiction. Pickard herself begins her explanation by pointing out that there are many (often good) reasons to choose to engage in addicted behaviors such as drug use, and so there is “no puzzle at all with respect to [the question of why addicts engage in their addicted behavior in the first place].”²⁹⁴ Hence, what may *appear* to outsiders as puzzling behavior is, quite often, not puzzling at all.

Pickard acknowledges that a puzzle does seem to remain in cases where the costs of the continued behavior exceed the benefits like those she alludes to (social acceptance, pain relief, and so on). Still, before addressing these cases (which she seems to think are the exception), Pickard highlights the fact that what does and does not count as a cost can vary widely from person to person. This is because factors that affect this determination, like personal values, socioeconomic status, cultural norms, social and political institutions, and the like, are highly variable from one addict to the next. Hence, Pickard again further constricts the boundary circumscribing the cases of addiction that remain puzzling on her view.

²⁹² If this seems like an exaggeration, it is not. This is a standard characterization of addiction on the prominent brain disease model of addiction found most often in medical, experimental, and clinical contexts. This is partly evidenced by the authors’ descriptions themselves as well as the fact that the standard responses from prominent anti-brain disease theorists is to appeal to evidence that addicts maintain some ability for control and choice. This implies (rightly, I think) that their target view can be expressed by a conditional: if addiction is a brain disease, then addicts cannot exercise control or choice (with respect to their addicted behavior). The opponents of this view simply *modus tollens* the conditional (usually implicitly). To be sure, this does not imply that the brain disease model *requires* this understanding of addiction. The point is just that, as far as the literature goes, this is the standard version on offer.

²⁹³ Pickard (2019, p. 13).

²⁹⁴ Pickard (2019, p. 13).

In the end, Pickard describes five influences on addicts' behaviors that might explain why they continue to *choose* this behavior despite overall negative consequences:²⁹⁵

- (i) **Self-hatred and self-harm:** sometimes addicts *want* and *intend* to cause themselves overall harm (for instance, due to an abusive upbringing, co-occurring mental health issues, and so on), and so they lack the concern or care for themselves that would make the (apparently) negative consequences an incentive to stop;
- (ii) **The least bad option:** sometimes addicts' circumstances (such as low socioeconomic status, poor mental and physical health, lack of opportunity) are such that the negative consequences of the addicted behavior end up being the best (or least bad) option for minimizing their suffering, compared to the perceived risk of abstinence;
- (iii) **Temporally myopic decision-making:** sometimes addicts shift their preferences (away from abstinence) because when the choice to engage in their addicted behavior is imminent, they perceive the immediate rewards to outweigh the (more long-term) costs, and they do so even when they maintain the opposite preferences in circumstances where the choice is not imminent;
- (iv) **Denial:** sometimes addicts deny (either explicitly or through mechanisms like motivated bias or self-deception) that their addicted behavior has (or is the cause of) the purported negative consequences, and so they do not perform the cost-benefit analysis for engaging in *that* behavior with *those* consequences in mind;
- (v) **Self-identity:** sometimes addicts see their addiction (and so their addicted behavior) as part of who they are, and so they either: (a) engage in that behavior habitually, accepting the negative consequences as part of the deal (or not thinking of them much at all); or (b) find meaning, community, and value in their addict identity, changing their perception of the purportedly negative consequences (for instance, they may identify with and so value them, they may fear losing their identity more, and so on).

Pickard's explanations of the puzzle of addiction are interesting and informative. The explanation from proponents of the brain disease model is clean and simple. However, both are problematic, and highlighting their problems will serve to illustrate why the dispositionalist account can best explain the puzzle of addiction.

4.2.1.1 The Brain Disease Model's Explanation of the Puzzle

First, as noted, the dispositionalist account keeps much of the cleanliness and simplicity of the brain disease model's explanation but does not inherit the problematic aspects of its claims about

²⁹⁵ The following are summarized from Pickard's discussion (Pickard, 2019, pp. 14–19).

compulsion or disease. What makes its explanation attractive is that the puzzle is quickly solved by introducing an inability to control oneself. Why do addicts continue in their addicted behavior even when it is harmful? They are compelled by a brain disease (says that account), and so they cannot help continuing in that behavior. On the account I defend, addiction impairs control in such a way that addicts are *systematically* disposed to fail to control their desires to engage in their addicted behavior. Hence, we can still get a clean and simple explanation for many of the puzzling cases by appeal to impaired control. The addicts in these cases would normally stop, or work towards stopping, were they not addicted (this is why they can sincerely express regret and a competing desire to stop). However, their systematic disposition to fail to manage their desires makes this exceedingly difficult (such that they are regularly unsuccessful). To manage some behavior (or to cease it entirely) is to control that behavior, and this entails controlling (to some significant degree, at least) the desires that drive that behavior.²⁹⁶ Thus, puzzling cases where an addict continues in their behavior despite unwanted harm is precisely what the dispositionalist account defended here would predict.

However, this account also avoids the problematic aspects of the brain disease model's explanation. This is because it does not require that addicts are compelled in the sense that they cannot do otherwise than engage in the addicted behavior when presented with the opportunity to do so. This is where choice model proponents like Pickard and others resist the brain disease model's simple explanation: the agency-undermining, literally irresistible sense of 'compulsion' is *too* simple. The systematic disposition to fail to control one's desires is triggered by, roughly, certain opportunities to engage in that behavior together with attempts to control it. However, the realization of this disposition (failed attempts of control) need not be deterministic. It will be sufficiently likely to occur in the appropriate conditions, but not necessarily *guaranteed*. Finally, the account I defend here makes no claims about disease in defining 'addiction'. The account is *consistent* with particular instances of

²⁹⁶ By 'manage their desires', I am alluding to the discussion in the preceding chapter about indirect control over desires.

addiction (even *every* instance) being a disease. But again, the resistance from the choice model proponents to such disease claims ultimately comes to the thought that calling addiction a ‘brain disease’ entails lack of agency, choice, or an ability to resist engaging in the behavior. Conceptually linking ‘disease’ and ‘compulsion’ in this way is a mistake, and the dispositionalist account I defend avoids this worry in any case.²⁹⁷

4.2.1.2 Pickard’s Explanation of the Puzzle

Second, regarding Pickard’s choice-model-based explanations of the puzzle of addiction, there are three important problems. The first is that, as noted, in moving away from the appeal to impaired control, Pickard is forced into trying to mostly explain away the puzzle of addiction. She points to the various reasons for drug use that would explain why the puzzle is only apparent from an outside perspective. If we were to get in the heads of the addict, she suggests, we would see that they are typically just choosing to engage in the behavior because they simply do not see the cost-benefit analysis in the way we would expect (having overall negative or harmful consequences). This is why she says that it is often more accurate to describe the phenomenon as a “puzzle of choice,” and also that this sort of puzzle is often really “no puzzle at all.” This is a counterintuitive feature of a choice model like Pickard’s since surely there are some puzzling cases that remain to be explained; after all, this would best explain why the puzzle of addiction has been such a predominant target of study in addiction research for so long. The account defended here avoids this counterintuitive result.

Even if we wanted to *mostly* explain away the puzzle, we could not do so entirely. Some addicts do engage in their behavior despite harmful consequences *that they recognize are harmful and worth avoiding*.

²⁹⁷ See Wakefield (2017a) for a discussion of the fact that many addiction researchers make such mistaken entailments in the disease vs. choice debate – on both sides. Wakefield also defends a version of the brain disease model that is much more nuanced in this respect. He requires impaired control, but he does not invoke such an implausible account of compulsion. However, it should be noted that he is a rare exception in the literature in terms of his care and nuance with the brain disease model. As with other choice model proponents like Pickard, Lewis, Levy, Heyman, and others, I think it is important to direct attention to and address the version of the brain disease model that is most prominent and influential. This, unfortunately, is the one that invokes implausible views of compulsion.

That is, there *really is* a puzzle of addiction. Pickard seems to recognize this, offering five explanations for the remaining puzzle cases. This leads to the second problem with her account: each explanation ends up again trying to explain away the puzzle. Once again, however, this is the wrong result. Whether it is because of self-hatred, impoverished circumstances, temporally myopic decision-making, denial, or self-identity, each explanation Pickard gives ends up claiming that the addict *does not* see the consequences of their addicted behavior as overall negative or harmful. Hence, Pickard ends up in the same position of trying to *explain away* the puzzle, which the dispositionalist account need not do.²⁹⁸

The third problem for Pickard is that her explanations are all best interpreted as invoking some degree of loss of control. For instance, her discussion of denial involves the influence of motivated biases and self-deception and her discussion of temporally myopic decision-making involves preference shifting at the moment of choice. Each could count as an instance of impaired control. Someone's addiction might manifest in their desires clouding their deliberation, resulting in motivated reasoning or self-deception about what they think is best or how they view the relationship between their behavior and the harm they experience. It may also manifest in an inability to maintain focus on one's long-term goals, plans, or resolutions when the choice to engage in the behavior is imminent. Some researchers identify this sort of preference shift as the locus of both the impaired control and the disorder that, on their view, is the addiction.²⁹⁹

As previously discussed, impaired control over one's desires comes in different forms and is instantiated in different stages of the process of trying to modify those desires. That her explanations are best interpreted as still involving impaired control is a problem for Pickard because she is trying

²⁹⁸ To clarify a bit further, by 'explaining away' the puzzle I mean that Pickard has to end up saying that the phenomenon we thought was there is, in fact, not there. This is in contrast to explaining *why* the phenomenon occurs. The phenomenon – the puzzle of addiction – is that addicts continue in their addicted behavior despite (often severely) harmful or negative consequences *which they want to avoid*. Hence, *explaining away* the puzzle means that addicts do not behave in this way. *Explaining* the puzzle means offering an explanation for why they do.

²⁹⁹ See, for instance, Levy (2014, 2019).

to avoid an appeal to impaired control *in explaining the puzzle of addiction*. This problem bodes well for the dispositionalist account since it can utilize all of the evidence-based explanations that Pickard provides in explaining the puzzle of addiction, yet still maintain that what these explanations are getting at are varying degrees of varying types of impaired control.

Hence, the dispositionalist account of addiction can take what works from both the brain disease and choice models in explaining the puzzle of addiction. Perhaps some puzzling cases are only apparent and get explained away – after all, addicts retain some choice on my account. However, for those puzzling cases that remain, the systematic, non-determinate disposition towards impaired control over one’s desires best explains them. What is more, it can do so in conjunction with Pickard’s and others’ appeal to factors like self-identity, context, co-morbid mental health issues, and the like, that influence control.

4.2.2 Addiction Is Janus-Faced

The foregoing discussion points to another fact about addiction that must be explained by any good account: addicts often suffer impaired control while still maintaining some capacity for choice. This *Janus-faced* feature of addiction is missed by views on the two extreme ends of the spectrum regarding control and choice.

On the one hand, brain disease models claiming that addicts suffer from agency-undermining compulsion fail to capture the fact that addicts quite often seem to control their choices regarding the very behavior they are supposed to be addicted to engaging in. As noted, choice model proponents have been keen in pointing this out. Consider the following examples:

- (i) Addicted smokers can refrain from lighting up while on long flights.
- (ii) Most addicts age out of their addictions without treatment, and for very normal reasons such as concerns over family, values, finances, health, and so on.³⁰⁰

³⁰⁰ See Heyman (2009, 2013).

- (iii) Addicts reduce consumption when prices on their preferred substance are raised.³⁰¹
- (iv) In some studies, addicts choose small monetary rewards over their drug of choice.³⁰²
- (v) Addicts can be competent to consent to clinical trials involving their drug of choice.³⁰³
- (vi) Addicts will sometimes refrain from use in order to reduce their tolerance, resulting in increased psychoactive effects.³⁰⁴

On the other hand, the naïve moral model, holding that addicts suffer no impairment in control but only a defect in character, is undermined by the plethora of evidence (described in the previous chapter) that addiction impairs one’s capacities for control.³⁰⁵ Hence, neither extreme can sufficiently account for the Janus-faced nature of addiction. What I refer to as the ‘choice models’ fall somewhere in between, pushing back on claims of compulsion yet acknowledging the abnormal difficulty that addicts face in reigning in their behavior.

What choice models need – or rather, what they need to be more explicit about – is a way to explain the intermediate loss of control that fits into the middle ground they attempt to carve out in these debates. The dispositionalist account offers just that. The evidence brought against claims of compulsion and supporting some degree of choice, such as (i)-(vi) above, is consistent with an impaired *capacity* for control regarding their addicted behavior. More to the point, an impairment in general capacity still allows for *exercises* of control in certain (even many) instances regarding the relevant behavior. Furthermore, it even allows for normal levels of control when the capacity is assessed against behaviors that the individual is *not* addicted to. In other words, not only can a heroin

³⁰¹ See Xu & Chaloupka (2011).

³⁰² These are called ‘contingency management interventions’. See Ainscough et al. (2017) for review.

³⁰³ The plausible assumption here is that compulsion undermines genuine consent. See Henden (2013b). See Uusitalo & Broers (2015) for a critical response, and Henden (2016b) for a reply to their response.

³⁰⁴ Discussed in Heyman (2009). It is perhaps difficult to say whether this intentional abstinence is actually evidence of control. Perhaps the decision to remain temporarily abstinent is itself uncontrolled since the desire is strong and the addict believes this is the only way to satisfactorily fulfill it. Of course, it is entirely consistent with this evidence that it *is* an instance of controlled choice. Plausibly, both interpretations are true since it is likely a mix of these.

³⁰⁵ Keep in mind that ‘moral model’ is basically a technical term. It does not refer to any theory of addiction that says addicts act wrongly, since this is consistent with some views of impaired control. One can act wrongly and not be blameworthy, so acting wrongly without control need not worry us. The views I am referring to by the use of ‘moral model’ are those that deny any loss of control in addiction – an addict chooses to use like a non-addict chooses a movie.

addict sometimes exercise control over their heroin-related desires, but they can also maintain a normal capacity for control regarding their other non-heroin-related desires, such as for shopping, playing basketball, and so forth. This is precisely what the dispositionalist account allows, and hence, it predicts (and explains) the Janus-faced nature of addiction.

4.2.3 Addiction Is Context-Sensitive

A third fact about addiction is that it is sensitive to context. For instance, environmental factors like opportunity and socioeconomic status can play an important role in influencing whether, how, and to what extent an addiction either develops, continues to exist, or is realized. As Hanna Pickard has recently put it:

...the minority of people who do not spontaneously recover from addiction typically come from underprivileged backgrounds of severe adversity and limited socioeconomic opportunity and suffer from a range of mental disorders in addition to substance use disorder. Addiction flourishes in conditions of poverty, isolation, humiliation, pain, and hopelessness of the sort many of us who are more privileged will never know. In such circumstances, drugs may be the only thing that brings any relief from suffering and despair.³⁰⁶

Consider also two well-known examples of addiction – or, at least, addiction-like behavior – being highly dependent on context: Bruce Alexander’s Rat Park study and the widespread recovery of opiate-addicted soldiers upon returning home from the Vietnam War.³⁰⁷ In the Rat Park experiment, rats that were kept in environments containing social elements such as an exercise wheel and fellow rats to play and mate with, as well as a choice between a drug-laced and non-drug-laced drinking supply, were far less likely to exhibit addiction-like behaviors as compared to rats kept in individual cages without these features. In the studies of returning Vietnam soldiers, it was found that surprisingly high numbers of soldiers who were by all appearances addicted to opiates while deployed and fighting in the war had suddenly stopped using upon returning home – without receiving any treatment.

³⁰⁶ Pickard (2020, p. 42). Note again the way in which this seems to be explaining *away* the puzzle of addiction.

³⁰⁷ See Alexander (2010) and Hall & Weier (2017), respectively.

There are a couple of things the dispositionalist account can say about all of this. First, regarding Pickard's emphasis on environmental influences of addiction, it should be unsurprising that the development and realization of our dispositions are influenced by our environment. Dispositions have triggering conditions – types of circumstances that can help to realize the disposition when it is embedded in them. Coupled with the fact that dispositions vary in strength – their likelihood of being realized when triggered – this means that some types of circumstances will be more likely than others to trigger a given disposition. For instance, my ability to make tennis serves is less likely to be realized when I am on the side of the court exposed to the sun. A person with pollen allergies can avoid flare-ups by moving to Utah or Colorado,³⁰⁸ and I may exhibit allergy-like reactions to pollen if it were shoved into my nose. Dispositions are sensitive to context in this sense because it is when they are (or are not) in certain circumstances, with certain features, that they can be triggered according to their likelihood. Something similar is also true of their development. It is when their potential bearers (with all of their other properties) are (or are not) in certain circumstances, with certain features, that their physical makeup can be modified in the right way such that certain dispositions will (or will not) begin to develop. Pickard is simply highlighting the evidence for some of these relevant triggering and developmental conditions for addiction: poverty, denial, self-identity, co-morbidities, and so on.

Second, regarding the Rat Park and Vietnam Vets phenomena, the dispositionalist account gives us a reasonable explanation to capture what is going on. The most plausible explanation seems to be that not all of the Vietnam Vets were actually addicted; and similarly, not all of the rats in (non-Rat Park) isolation-style animal models of addiction were addicted (if any were). These contexts are akin to dropping bowling balls from the Empire State Building. To call such bowling balls 'fragile' when they break would be as mistake, just as it would be a mistake to think that such a context was a

³⁰⁸ Which are apparently two of the best places to live in U.S. if one suffers from allergies. See here: <https://www.rentcafe.com/blog/apartmentliving/lifestyle/best-places-to-live-with-allergies/>.

relevant triggering condition for testing fragility. The same is true of being in the heat of the Vietnam War or being isolated in a cage with nothing but cocaine-laced water. Anything or anyone is liable to break in such conditions.

On the dispositionalist account, addiction is fundamentally about what the bearer will do in all of the relevant possible circumstances rather than just one interesting circumstance. Hence, *even if these were relevant triggering conditions*, it simply would not entail that they were addicted. Moreover, the evidence that many subjects behave entirely differently in different contexts suggest that they are not relevant triggering conditions. Vietnam Vets were under extreme pressure and threat of death. Yet, when those who came home were placed back into societally relevant choice conditions, we see that many ended up exercising control over their desires (or perhaps lost those desires altogether). The much smaller majority who remained addicted were those who continued to fail to control their desires even in much more ordinary circumstances. The dispositionalist account can explain this and would count only those who are *systematically* disposed to fail to control their desires as addicts, rather than anyone who fails to do so in some particularly extreme circumstances.³⁰⁹

As we can see, the contextual sensitivity of addiction and the possibility of exhibiting addiction-like behavior in particular contexts fits well within the dispositionalist framework.

4.2.4 Addiction Is Sometimes Indeterminate

A fourth fact to explain is that addiction seems to fall on a continuum, where some are clearly not addicted, others are clearly addicted, and still others (perhaps many others) are neither clearly addicted nor clearly not addicted. In other words, addiction appears to involve a grey area of vague or

³⁰⁹ An additional feature of my account is that, if we are so inclined, we can contextualize our ‘addiction’ ascriptions by restricting the reference class in certain ways, which ultimately modifies which circumstances count as triggering conditions. For instance, if we restrict our reference class to *the soldiers in the Vietnam War* (or battlefield contexts more generally), we can say that those soldiers were *addicted-in-the-war*. Not every single soldier uses drugs or exhibits signs of addicted use, and hence, even within that restricted context there are differences in terms of controlling one’s desires. Importantly, this is not to change what addiction is and so is distinct from their being addicted *per se*. It is simply a way of speaking that illustrates how my account could help us to describe what is going.

indeterminate cases. This is reflected in the DSM-V's spectrum-style diagnostic criteria allowing for mild, moderate, and severe levels of *substance use disorder* based on the number of criteria met.

The dispositionalist account not only allows for grey area cases, but it predicts that such cases would exist. Recall that addiction is about being *systematically* disposed to fail to control one's desires, and this involves being sufficiently strongly disposed in a sufficient portion of the relevant triggering conditions. This wording might seem to suggest a binary account, wherein some degree n of strength in some amount m of the relevant conditions is what tips the scales from non-addict to addict. However, the account is not committed to postulating the existence of this kind of threshold. Perhaps there is a hard cutoff and the grey area cases amount to epistemic vagueness. But perhaps there is not, and some cases of addiction are truly indeterminate or metaphysically vague. In either case, we want an explanation for why there is (or appears to be) this spectrum of addiction with plenty of hard cases in the middle. On the account I defend, this is straightforwardly due to the dispositional nature of addiction, and the varying degrees of strength, opportunity, and systematicity of that disposition.

4.2.5 Addiction Takes Many Forms

A fifth fact about addiction is that it seems to take on a wide variety of forms, most notably in the different possible objects of addiction.³¹⁰ A growing body of evidence supports the existence of so-called 'behavioral' or 'process' addictions, such as gambling, internet, sex, or exercise addiction.³¹¹ Moreover, television shows like "My Strange Addiction" follow subjects purportedly addicted to such things as eating drywall, drinking white out, and pulling hair out of drains, and there

³¹⁰ Keep in mind that by 'objects of addiction' I do not mean material substances. On my view, all addictions are to certain types of behavior, such as ingesting, seeking, and procuring substances like heroin.

³¹¹ See Yao & Potenza (2015), Shaw & Black (2008), Karila et al. (2014), and Berczik et al. (2012), respectively.

are even reports of water addiction.³¹² This evidence needs to be explained (or explained away), and it raises the question of whether people could be addicted to anything.

Fortunately, the account defended here does not require us to explain this evidence away. Of course, it is possible that *some* of these cases do not involve genuine addictions. However, just as we saw with the Vietnam Vets and Rat Park examples, the dispositionalist account is fully capable of capturing the possibility that particular (sometimes abnormal) contexts can produce addiction-like behaviors. That said, a benefit of this account is that it can fully accommodate the reasonable and more likely explanation that at least some of these cases involve genuine addictions.

It might sound odd to say that someone is (or even could be) addicted to water, but this is not yet a reason to deny the possibility. All the dispositionalist account defended here needs is that this person has desires for certain types of water-related behaviors (ingesting it, seeking it out, keeping it on hand, and so on), and that they are systematically disposed to fail to control them. Moreover, given the reward-based view of desires, we can explain how this differs from ordinary, healthy urges to drink or procure water when thirsty (autonomic impulses are not the same as the reward-based desires involved in addiction). If one's environment is set up just right and a certain sort of history unfolds, an individual can certainly develop strong (reward-based) desires for water (or just about anything, really). What is more, the account I defend requires neither that addiction is diseased nor that it is harmful. Hence, the possibility that some of these addictions are not yet instances of some dysfunction and not yet leading to any harm is neither here nor there. Contrast this with any view that requires harm or disease as a necessary condition of addiction – and there are plenty to choose from considering almost everyone takes at least one of these conditions to be a definitional feature.

³¹² See an episode list for “My Strange Addiction” here: https://en.wikipedia.org/wiki/My_Strange_Addiction. See Hutcheon & Bevilacqua (2010) and Edelstein (1973) for water addiction.

The brain disease views also have an additional problem accommodating behavioral addictions since they are often based in the potential for the chemical properties of commonly addictive substances to hijack (as they call it) the proper functioning of the certain brain systems. Heroin, for instance, directly triggers a reward signal independently of the reward system's normal feedback loop – roughly, it makes your brain say, “Great, better than expected!” no matter what.³¹³ If that is the dysfunction of addiction, as many of these theories suggest, they will have a hard time explaining what is going on in behavioral addictions that do not involve ingesting chemicals that directly interact with our reward system. I am not the first to point out this puzzle for brain disease views.³¹⁴ The difficulty the brain disease view has with bringing substance and behavioral addictions under the same account is also reflected in the fact that only two behavioral addictions are accounted for between both the ICD-11 and the DSM-V (gambling disorder and internet gaming disorder, respectively).³¹⁵ While empirical work is shedding a light on the role of the reward system in behavioral addictions, the upshot here is that the dispositionalist account is not threatened in the slightest by the possibility – and likely fact – that addiction comes in many forms and is manifested in many ways.

4.2.6 ‘Addiction’ Is Polysemous

A sixth and somewhat different sort of fact about addiction is that its label, ‘addiction’, is used in different ways in different contexts. Unlike the contextual sensitivity of addiction itself, though, this is not exactly a fact about addiction *per se*. Nonetheless, I want to argue that it is worth capturing. First, if there is reason to think that some non-trivial amount of the common usage of ‘addiction’ is sufficiently similar in meaning despite some variation, then an account of addiction that captures the

³¹³ Recall that this would be a *positive learning signal* on our reward theory of desire, and not necessarily pleasure.

³¹⁴ See, for instance, Lewis (2015).

³¹⁵ See the ICD-11 (World Health Organization, 2020) and the DSM-V (American Psychiatric Association, 2013).

underlying meaning shared across different uses is preferable. Second, I contend that there is reason to think this is true of the common uses of ‘addiction’.

‘Addiction’ is a term like ‘fight’. Some fights are fist fights. However, when Tyson says, “I fought with my wife for hours last night,” he does not (necessarily) mean that they were in a physical altercation. This is because some fights are (non-physical) spousal arguments. Hence, Tyson speaks sincerely and meaningfully when he says such things. There are also dogfights (the aerial kind), patients fighting for their lives, and perhaps still more uses of ‘fight’. There are obvious differences between these different uses of ‘fight’, but just as obviously there is something that seems to unify them (some type of struggle perhaps). My claim is that ‘addiction’ is like this.

Talk about addiction will vary depending on the context one is in, and there does seem to be something like different kinds of addictions in the way that there are different kinds of fights. We speak of addictions in the ordinary sense, as in someone’s being addicted to alcohol or heroin. But we also speak of our children being addicted to video games, television, or cell phones, and of our friends being addicted to chocolate or Netflix. We might even recommend a television series to a friend by calling it ‘bingeworthy’ and even issuing a warning that it is ‘so addictive’, which is (perhaps counterintuitively) meant to be a selling point. Companies promoting cell phone games and apps often explicitly advertise their product by calling it ‘addictive’, implying that this is a good thing. Artists across generations and genres sing of love as an addiction.³¹⁶ We speak of addiction as a social phenomenon, a medical phenomenon, a biological phenomenon, a historical phenomenon, a moral phenomenon, a phenomenon of agency, and so on.

³¹⁶ To name just a few: Avicii, “Addicted to You”; BeBe and CeCe Winans, “Addictive Love”; Billie Holiday, “You Go to My Head”; Diamond Rio, “Beautiful Mess”; The Gun Club, “She’s Like Heroin to Me”; Huey Lewis and the News, “I Want a New Drug”; Johnny Gill, “A Cute Sweet Love Addiction”; Ke\$ha, “Your Love Is My Drug”; The La’s, “There She Goes”; Lucinda Williams, “Essence”; Robert Palmer, “Addicted to Love”; The Weeknd, “Can’t Feel My Face”; XTC, “You’re My Drug”; and Zac Brown Band, “Beautiful Drug.”

While there are differences between these various uses of ‘addiction’, and between the phenomena that they pick out, there still seems to be something of a unifying nature that each is getting at. My contention is that what these uses are getting at, albeit in different ways, is the idea that the individual is experiencing (or will experience) some strong motivation towards the relevant behavior (or object) such that they will have an abnormally difficult time resisting it. By ‘abnormal’ I do not mean to imply disease or dysfunction but am only pointing to the idea that the difficulty is noteworthy in some (perhaps vague) way – such is the nature of colloquial discourse. The main point, though, is that if this is right, then the account of addiction defended here can make sense of these varied uses of the term. Moreover, it can do so without invoking any conception of disease or harm, which is likely not meant to be invoked by common users of the term in many cases. For instance, the bingeworthiness of Netflix shows and the addictive nature of the Bejeweled app are likely not meant to refer to any kind of harm, dysfunction, or disease.

Insofar as an account best captures the varied common usage of the term ‘addiction’, it is preferable to its competitors, all else equal. The dispositionalist account is preferable in this sense.

4.2.7 Addiction as It Can Be

The final facts about addiction to be captured by any good account pertain to metaphysically possible cases of addiction. Above I motivated the importance of this criteria for getting at the nature of addiction. Here I identify some possible cases of addiction that the dispositionalist account captures. Other extant accounts of addiction either cannot capture some of these cases or it is presently unclear how they could (for instance, such accounts require harm, a history, actual use, or other factor that is absent). When it is unclear, the burden is on the defenders of those accounts to explain how the case(s) could be captured.

4.2.7.1 Harmless Addictions

Consider first the possibility of harmless addictions. One thing to note up front is that it is unclear whether this is a *mere* possibility. As noted in the preceding section, it is certainly common to speak about addiction in a way that does not imply anything about harm occurring. But even setting this aside, it is at least metaphysically possible that an individual could be addicted without suffering any harmful consequences.

Suppose Ron is an alcoholic. He experiences strong desires to drink and is systematically unable to control them when he tries. However, Ron has no friends or family to be harmed by his drinking. Imagine further that a treatment has been developed that eliminates any alcohol withdrawal symptoms and other negative health effects from alcohol consumption, so Ron himself is also not harmed by the drinking. Ron is also independently wealthy, and so does not need to work. The case can be somewhat tricky since, due to the dominance of the medical model in academic and popular media, it might be hard to imagine an alcoholic who is not in some way harmed. Still, keep in mind that the point is about what is metaphysically possible and not what is common for us to associate with alcoholics. The upshot here is that an account that requires harm has to maintain that such cases are impossible. The account I defend, on the other hand, can accommodate the possibility of harmless addictions. The individual need only have the relevant disposition; the fact that some typical effects or manifestations of that disposition are not actualized in this case is of no consequence.

4.2.7.2 Instant Addicts

Consider next the possibility of instant addicts. For instance, suppose God either creates an addict from nothing or instantly turns a non-addict into an addict. Accounts that require some particular type of history – or any history at all, for that matter – cannot accommodate such possibilities. They would have to maintain that either God could not create an addict without a history

or that whatever is created is not an addict.³¹⁷ In other words, they would have to explain away this possibility, which seems much less preferable than an account that can straightforwardly capture what is going on. The dispositional account can do just this.

The simple explanation is that God created a being with certain dispositions (and one important disposition in particular) or, alternatively, that God modified an existing being's dispositions. Consider the changes that occur when I go from being very poor at hitting tennis serves at t_1 to having acquired at t_2 the systematic disposition to make tennis serves when I try. Ordinarily there is a series of processes that unfold between t_1 and t_2 that constitute the development of the relevant disposition(s), and which culminate (at t_2) in my acquiring the new disposition *tennis serving expertise*. In the case of God creating an instant addict, the difference is just that the process of change gets reduced to a single instant (or in the case of creation from nothing, God simply brings the t_2 state into existence at some instant). By analogy, consider God zapping a non-fragile vase so that it instantly becomes fragile (or creates a fragile vase from nothing). Fragility is about being disposed in a certain kind of way and so here, too, God is simply zapping the disposition of fragility right into the vase. Addiction, being a disposition, would work the same way.

The dispositionalist account holds that addiction is a certain type of disposition but is (or at least can be) silent about how or why that disposition came to inhere in the individual that bears it.³¹⁸

4.2.7.3 Addictions that Never Show Themselves

Consider next the possibility of addictions that are never realized. It is possible that an addict could exist without their actually engaging in any addicted behavior. For instance, consider the first few minutes of the instant addict's life. Perhaps God creates him on a desert island with no object of

³¹⁷ I will assume that God creates an *earthly organism* that can be addicted. A person is the easiest example.

³¹⁸ Note that this does not make the causal story uninteresting or irrelevant. It only makes it non-essential to *being an addict*. That is, two people with the relevant disposition are both addicts even if one came to acquire it through an ordinary history of, for instance, substance use while the other came to acquire it through the more extraordinary power of God.

addiction (or its relevant cues) to trigger his desires or subsequent attempts to control them. Alternatively, consider simply placing an addict on a desert island. This person would still be an addict in the same sense that a fragile vase that is bubble wrapped and thrown into a vault, never to be touched again, is still fragile.

The explanation is that *fragility* and *addiction* are dispositions, and so they explain what their bearers are *prone* to do in the right sorts of circumstances, and not necessarily what they are doing now. Indeed, we bubble wrap the vase *precisely because* it is fragile, whether or not it has broken yet. An addict actually engaging in their addicted behavior is not a necessary condition of the presence of an addiction for the very same reason that actual breaking is not a necessary condition for the presence of fragility. Hence, the dispositionalist account easily explains how unrealized addictions are possible, whereas accounts that appeal to actual behaviors being performed (for instance, those requiring *continued use despite harmful consequences* or *repeated failed attempts to quit*) cannot capture this.³¹⁹

This can also explain what is going on with many cases of relapse, even years after the addicted behavior has ceased and dependence is long gone. Of course, not every case of relapse is explained by a dormant disposition since it is possible that the addiction disposition can go away and then be re-acquired again later. However, it is certainly possible that for some strong addictions, the disposition becomes dormant and does not actually go out of existence. This might be because the treatment provided only masked the disposition or because no treatment was provided, and the individual only moved themselves away from their ordinary triggering situations. Whatever the case may be, the dispositionalist account I defend easily explains these cases of relapse. The disposition was just not being realized until it found itself in the right (or wrong) circumstances again.³²⁰

³¹⁹ Volkow (2014) is an example. Such views also cannot capture instant addicts since there is no addicted behavior at first.

³²⁰ There might be some cases where the facts underdetermine whether the disposition was masked or absent.

4.2.7.4 Alien Addictions

Finally, consider the possibility that aliens can be addicted, who may have brains with a different physical makeup than we do. As with the other possibilities, the answer here will again be quite simple: the aliens would bear the systematic disposition to fail to control their desires for certain types of behaviors. A few points are important to keep in mind. First, I favor a reward theory of desire that is grounded in reward learning, which is tied to an evolutionary history of Earthly organisms like humans and other animals. Nonetheless, a core part of the dispositionalist account is that there is some motivational state of the organism that drives thoughts, feelings, and behavior, and which can modify the organism's dispositions according to whether and how they act on those motivations, and the subsequent effects of doing (or not doing) so. Humans have brain systems like the mesolimbic dopamine system underneath all of this, but it seems to me that this is simply a contingency of our evolutionary history. So long as the alien can desire (or be motivated towards) some type of behavior, and has some capacity for control, then the dispositionalist account can accommodate this.

Second, and relatedly, the dispositionalist account I defend is control-neutral to a significant degree. So, as long as the alien has some typical means of controlling their desires and behaviors, then there will likely be degrees of control that they can exercise. For instance, perhaps there are reasons they are typically sensitive to when healthy or rational, perhaps they have agent-causal powers, and so on. If so, then an alien addict would bear the systematic disposition to suffer impairment of their control over the aforementioned motivational states.³²¹

Contrast this with the brain disease model that appeals to malfunctioning of particular areas of the human (or mammalian) brain. This focus on the material basis of the disposition is what can

³²¹ If the reader is skeptical of discussing alien addictions (despite my initial caveats on metaphysical possibilities above), then it seems that we could change the example to be about humans being addicted to newly discovered drugs that affect the brain in completely different ways, or about newly discovered species on our planet. All we need is that there is some sense of motivation and some sense of capacity for control over those motivations.

cause this type of view problems in trying to accommodate the possibility of addicts that may have wildly different matter, parts, and processes underlying their mental states and behaviors. Moreover, contrast the dispositionalist account's ability to capture alien addictions with an environment-based view of addiction like the ones Pickard, Alexander, and biopsychosocial model proponents defend. The possibility of alien addictions either cannot be captured by such accounts or it reveals that these accounts were not really telling us about the nature of addiction after all. Instead, they are telling us about what sorts of factors commonly *influence* the development and manifestation of human addictions. This is important work; still, it is distinct from *what addiction is* and so is not suited for capturing metaphysical possibilities like the one under consideration. The account defended here gets at the dispositional nature of addiction, and therefore, *can* accommodate these possible cases.

4.2.8 The Dispositionalist Account Is Explanatorily Powerful

Let us take stock of the preceding section. A good account of addiction ought to be able to explain the key phenomena related to it. At the very least, it should be consistent with the best explanations of such phenomena. I argued that the dispositionalist account meets these requirements, often quite easily, when it comes to such facts as: the puzzle of addiction; the Janus-faced and context-sensitive nature of addiction; the existence of indeterminate cases of addiction; addiction taking on many different forms; the varied ordinary use of the term 'addiction'; and four metaphysically possible cases of addiction.

Let us now turn to capturing other extant accounts in the literature.

4.3 The Dispositionalist Account of Addiction Captures Extant Views

This section explains how the dispositionalist account moves towards substantive unification by capturing extant accounts in the literature. First, I argue that, when charitably interpreted, most accounts on offer are either implicitly or explicitly describing addiction as a disposition. Hence, the

account I defend captures what most others in the literature are already getting at. I also discuss why this does not make my account trivial or obsolete. Second, building on this first argument, I explain how any plausible view in the literature can be accommodated into the dispositionalist framework. Finally, I close the section with an example of how my account can expose where disagreements between accounts are merely verbal, and thus help to avoid such unhelpful exchanges.

4.3.1 A Disposition by Any Other Name

Most everyone in the addiction literature seems to be talking about the same thing. I do not mean this in the trivial sense that they are all talking about what they take addiction to be. I mean that virtually everyone who offers some account or characterization of the nature of addiction can be understood as identifying addiction with the same kind of thing in the world, namely, a disposition of some sort. This should be surprising to anyone with minimal exposure to the addiction literature (whether academic or popular), since it is easy to see that there are apparently an uncountable number of accounts on offer. Moreover, it is even easier to see that the literature (again, whether academic or popular) serves as a battleground for competing views of the nature of addiction. This is likely why the metaphor of the blind men and the elephant is so popular in discussions of addiction.³²² In other words, anyone who claimed that virtually everyone who writes on addiction seems to agree at a fundamental level about what addiction is would have some explaining to do. This is a claim I defend, and so let the explanation begin.

4.3.1.1 Someone Had to Say It

The first bit of explaining needs to combat an intuitive reaction to the foregoing. The reaction is that my thesis (perhaps my whole account) must be trivial or obsolete if, as I claim, it defends a view

³²² Many examples of its use exist (du Plessis, 2014a; Henderson, 2000; Kalant, 1989; Peele, 1987).

of addiction that everyone else already has. If everyone already thinks that addiction is a disposition, what good is my account?

The first thing to say is that, at best, most accounts only *implicitly* view addiction as a disposition. Thus, it is not as clear as one might think that the authors of these accounts would explicitly endorse the dispositionalist framework. Second, setting aside what any or all of these authors would agree to, the bottom line is that virtually no one is actually saying that addiction is a disposition, let alone laying out this sort of view in any robust kind of way. I refer the reader to Nick Heather's quote from Chapter 1, wherein he lamented that most addiction researchers simply assume, for whatever reason, that everyone knows what they are talking about when they talk about addiction. Third, even if everyone would agree that addiction is a disposition, and even if everyone were explicitly saying this, it remains true that no one in the literature has worked out in any detail what this would mean. There is a dearth of accounts focusing on the metaphysics of addiction. Of course, discussions about whether addiction is a disease, whether addicts lack control or autonomy, and the like are plentiful, which *can* reasonably fall within the realm of metaphysics.³²³ However, what is lacking is a fuller account of what it would mean for addiction to be a dispositional property, both metaphysically and in terms of how this would fit into the extant literature (including the science of addiction).

Thus, the difference (and novelty) in the account defended here is that it gets at the dispositional nature of addiction *explicitly*, and more importantly, it provides a *robust* account of what this disposition and its various features come to metaphysically. To be sure, a handful of accounts do speak in more explicitly dispositionalist terms. The difference, though, is in the fact that even these

³²³ I say 'can' because many such discussions occur within non-philosophical contexts, such as the clinical literature. Thus, it is usually a stretch to call such discussions 'metaphysical' since these are usually not philosophers, let alone metaphysicians, and it is unclear whether they would even welcome such an ascription. Their interests are typically medical and practical, not metaphysical.

accounts fail to fill out the metaphysical details. In the next two subsections, I provide evidence that most authors are speaking about a disposition when they speak about addiction.

4.3.1.2 Some Accounts Are More Explicitly Dispositionalist

I begin with some clearer cases of authors speaking in dispositionalist terms about addiction. Consider first Walter Sinnott-Armstrong and Hanna Pickard's account of addiction. They define 'addiction' as "a strong and habitual want that significantly reduces control and leads to significant harm."³²⁴ This view is plausibly dispositionalist since it is unclear how else to understand the use of 'habitual'. The addict, on their view, seems to be disposed to undergo strong wants with some regularity (perhaps in predictable circumstances) that impairs their control. But looking more closely at their understanding of control more clearly reveals that they are working within a dispositionalist framework. They propose a want-based account of control that is analyzed conditionally: if someone wants overall to perform (or not perform) a type of action, then usually they do it (or do not do it).³²⁵ To provide a conditional analysis of control is to say that control is dispositional in nature – a *capacity* that can be exercised to greater and lesser degrees in various circumstances. Finally, here are the authors defending the claim that their account of addiction can also accommodate willing addicts:

Notice that this definition applies equally to...those "willing addicts" who endorse their addiction and never try to control their use (Frankfurt 1971). What makes "willing addicts" willing is that they want overall to use drugs; so they do what they want overall to do when they use. Nonetheless, *if ex hypothesis* they became "unwilling" and no longer wanted to use...then they would still use at least usually. This is what makes "willing addicts" addicts.³²⁶

Here we have another conditional analysis of what it means to have an addiction – it is about what the addict *would* do *were* they in certain conditions. Taken together, the foregoing is indicative of a dispositionalist account without the label.

³²⁴ Sinnott-Armstrong & Pickard (2013, p. 861).

³²⁵ See Sinnott-Armstrong & Pickard (2013, p. 856).

³²⁶ Sinnott-Armstrong & Pickard (2013, p. 859).

Consider next Bennett Foddy and Julian Savulescu’s account of addiction. These authors define ‘addiction’ as “a strong appetite,” and then go on to define ‘appetite’ as:

...a disposition that generates desires that are urgent, oriented toward some rewarding behavior, periodically recurring, often in predictable circumstances, sated temporarily by their fulfillment, and generally provide pleasure.³²⁷

That these authors have a dispositionalist account could not be any clearer. Of course, there are differences between our respective views, as well as between their view and Sinnott-Armstrong and Pickard’s view. Most notably, Foddy and Savulescu claim that addiction is not different in kind from ordinary *akratic* action and that, as their qualifier ‘strong’ indicates, the only difference is an amplification of either the strength of desires involved or the frequency of undergoing (and succumbing to) them. Still, we can see that they understand addiction as a disposition at its core.³²⁸

As a final example, consider an ontological account from Janna Hastings, Barry Smith, and colleagues, which aims to represent the nature of addiction – type of entity it is and its interrelations with other kinds of entities.³²⁹ These authors proposed adding **ADDICTION** to the Mental Functioning Ontology (MF) as a subclass of **MENTAL DISEASE**. The main point for our present purposes is that, in BFO-compliant ontologies like MF, **DISEASE** *is_a* **DISPOSITION**. Hence, by understanding addiction as a type of mental disease, these authors understand addiction in dispositionalist terms. Their paper is quite preliminary and, as I have explained, no worked out ontology of addiction yet exists. Nonetheless, their proposal of what kind of thing addiction is remains clear.

While some authors make explicit claims to the dispositional nature of addiction, this is the exception. Let us turn to more representative examples, wherein authors *implicitly* describing addiction as a disposition.

³²⁷ Foddy & Savulescu (2010b, p. 35).

³²⁸ What is more, there is still room for impaired control on their account. As many choice account proponents are, Foddy and Savulescu are more concerned with combatting what they take to be the implausible conceptions of control, compulsion, and agency in brain disease accounts claiming that addicts lack these.

³²⁹ See Hastings et. al (2012b).

4.3.1.3 Most Accounts Are Implicitly Dispositionalist

There are roughly two ways to interpret most other accounts of addiction: either the authors have a disposition in mind when speaking about addiction or they do not. The first interpretation would seem to make sense of the plethora of dispositionalist terms used to characterize addiction and to describe the phenomena surrounding it, such as ‘susceptibility’, ‘tendency’, ‘trigger’, ‘cue’, ‘stimulus’, ‘vulnerability’, ‘chronic condition’, ‘repeated failures’, ‘conditioned response’, and so on. Still, some authors might not be intending to describe a disposition. Perhaps they are not intending to do metaphysics at all. Perhaps they actually mean that addiction consists in actual behaviors rather than some disposition or tendency towards those behaviors. The second interpretation gets at this possibility. However, in these cases the dispositionalist version is likely the more *charitable* interpretation of the view since it avoids certain counterexamples, such as addicts losing their addiction when they are asleep or between uses. At the very least, though, the dispositionalist view is compatible with the core of such accounts. Hence, even when dispositions are not the *explicit* locus of addiction, most accounts in the literature are either implicitly describing a disposition, ought to be understood in this way, or can easily be adapted to the dispositionalist view. Let us turn to some examples.

Consider first how Nora Volkow, the *de facto* face of the brain disease model, has described addiction in various works. She calls it “a conditioned response [following exposure] to the drug and/or drug-related stimuli”³³⁰ that involves a pathology in “how the brain regulates (chooses) behavioral output in response to those stimuli.”³³¹ She also emphasizes the fact that “the enduring vulnerability to relapse is a primary feature of the addiction disorder.”³³² These statements are indicative of a dispositionalist account. To be sure, Volkow focuses on dispositions of *the brain* to

³³⁰ Volkow & Fowler (2000, p. 323).

³³¹ Kalivas & Volkow (2005, p. 1410).

³³² Kalivas & Volkow (2005, p. 1410).

respond to environmental stimuli and then produce behavioral outputs of the agent, whereas I take addiction to be a disposition of *agents*. Still, we both have dispositions in mind.

Consider next Robert West and Jamie Brown's account from the second edition of their book *Theory of Addiction*.³³³ Like Nick Heather, West and Brown note the rampant failure of authors writing on addiction to state clearly what they mean by their terms, including 'addiction'. Their proposal is meant to move towards a remedy to this problem as well. They define 'addiction' as follows:

...a chronic condition in which there is a repeated powerful motivation to engage in a rewarding behaviour, acquired as a result of engaging in that behaviour, that has significant potential for unintended harm.³³⁴

Their description of the condition as 'chronic', of the powerful motivation as 'repeated', and the unintended harm as a 'potential' consequence is indicative of a dispositionalist account. Moreover, their reference to addiction being acquired as a result of engaging in the behavior gives a nod to associative reward learning, which they embrace in later chapters. This is clearly consistent with a dispositionalist account, as is seen in my earlier discussion of the reward theory of desire.

Notably, West and Brown intentionally avoid the use of 'impaired control' for two reasons:

- (i) "this might imply specifically that the disorder involves a weakening of self-control rather than an increase in drive to engage in the addictive behaviour,"³³⁵ and,
- (ii) "the focus on impaired control rather than the force causing the impairment excludes the possibility of addictions with the potential to cause harm where no attempt was (yet) being made to exert control."³³⁶

This reveals that their account is not *explicitly* dispositionalist since, were a dispositionalist account fully embraced, these would not be worries. First, both impaired self-control capacities and abnormally strong desires can be involved in the manifestation of the disposition to fail to control one's desires

³³³ West & Brown (2013).

³³⁴ West & Brown (2013, pp. 4, 15–16).

³³⁵ Note that West and Brown are careful to not make *the disorder* into impaired control, and do not necessarily resist the notion that addiction requires impaired control. The latter claim would be hard to square with their discussion throughout the book. They seem intent on identifying the increased drive, motivation, or craving aspects of addiction with *the disorder* rather than claiming that this is all that addiction is. Hence, their view here is consistent with my claim that the definition need not include disease or disorder to allow that a given instance of addiction still is or involves a disease or disorder.

³³⁶ West & Brown (2013, pp. 4, 15).

or behavior.³³⁷ Second, understanding addiction as a disposition means, roughly, that we do not have to worry about the disposition *actually* being manifested. If the systematic disposition is present, then, roughly, it will be true that *were* the addict to attempt to control their desires in the relevant conditions, they *would* fail with sufficient regularity. Impaired control refers to a capacity rather than that capacity being exercised.

As a final set of examples, let us consider seven other authors' descriptions of addiction. In each case, my contention is that the author is either implicitly describing a disposition or the most plausible way to interpret the view is by supposing they are referring to an underlying addiction. The latter is, again, the simplest way to avoid easy counterexamples like addicts losing their addiction when they are asleep. Such objections could otherwise be leveled against views that identified addiction with the actual exercises of a capacity rather than the capacity (disposition) itself. The following is a list of quotes from extant views in the literature that, in my estimation, have dispositions of some kind in mind (emphases have been added for easy detection of the dispositionalist descriptions):

- (i) **George Ainslie:** "We naturally overvalue the imminent future, and learn *the ability* to maintain long-term plans only gradually and imperfectly. Everyone struggles with *bad habits*... Those of us who have avoided the named addictive diagnoses are nevertheless *apt to suffer from habitual overvaluation of the present moment*, as in *chronic procrastination*, overuse of credit, or unrealistic future time commitment. So the problem for the science of addiction is not an addict's *susceptibility to temptation*, but why she fails to use her culture's shared knowledge to counteract it in specific areas over part of her life;"³³⁸
- (ii) **Nick Heather:** "...addiction, defined as *repeated and continuing failures to refrain* from a specified behavior despite prior resolutions to do so;"³³⁹
- (iii) **Edmund Henden:** "The property of 'being addicted' refers to a *certain kind of relation* a person has, not to some isolated act of consumption, but *to a pattern of behavior, enacted on a regular basis in characteristic circumstances*, which the person finds extremely difficult to

³³⁷ As before, the idea here is that addiction is very much like a skill in its structure since it will likely always rest on further dispositions. This is analogous to claiming that it is possible for two expert putters to both have the systematic disposition to sink attempted putts, while each golfer's disposition has a slightly different arrangement of underlying dispositions. One might have better muscle control and hand-eye coordination but weaker vision and balance, for instance.

³³⁸ Ainslie (2019, p. 37).

³³⁹ Heather (2017a, p. 147).

override by intentional effort...in fact, *it is not clear it even makes sense to speak of one-off addictive actions*);”³⁴⁰

- (iv) **Gene Heyman:** “Addiction is... voluntary behavior that *predicts the persistence of activities* that from a global bookkeeping perspective (e.g., long-term) are irrational. That is, addiction is not compulsive drug use, but it also is not rational drug use.”³⁴¹ And elsewhere: “the degree to which an activity is voluntary is the *degree to which it systematically varies as a function of its consequences*”;³⁴²
- (v) **Neil Levy:** “Due to features of the neurobiology of addiction, finding themselves on the garden path *disposes [addicts] to undergoing a cascade of judgment-shifts*: first shifting toward judging that they ought to step down the garden path and subsequently to judging that they ought to consume... [Moreover,] the behaviour of lapsing addicts often appears to be controlled behaviour...*it is incentive-sensitive* in a very ordinary way.”³⁴³
- (vi) **Marc Lewis:** “So, what exactly is addiction? *It’s a habit* that grows and self-perpetuates relatively quickly, when we repeatedly pursue the same highly attractive goal. Or, in a phrase, *motivated repetition* that gives rise to *deep learning*...So addiction is not fundamentally different from other unfortunate directions in personality development: a *self-reinforcing habit* based on intense emotions, *encountered repeatedly*”;³⁴⁴
- (vii) **Jerome Wakefield:** “[A form] of *disruption of the desire/deliberation/choice system* that [is] not biologically designed to be preemptory...In such instances, reasonable trade-offs between desires cannot be seriously contemplated, important balancing considerations are disregarded or summarily dismissed, and one motive very disproportionately dominates the deliberative process...Clearly, *this disruption is a matter of degree once it passes over some fuzzy threshold* that distinguishes the depth of the processes involved in, say, transient weakness of will in having too much pie at dinner from opiate addiction that destroys a family. In principle, severity of addiction should be measured in terms of *the degree and durability of the constriction of the desire/deliberation/choice system*.”³⁴⁵

As we can see, accounts from across the board – choice models, brain disease models, *no difference* models, ontological models – all describe addiction in dispositionalist terms, whether explicitly or implicitly. Many such examples exist in the literature. We should find this unsurprising since, as I see it, the dispositionalist account is an entirely natural and intuitive way to think about common features of addiction: the sensitivity to triggers/cues; the vulnerability to relapse when recovering; the

³⁴⁰ Henden (2016a, p. 124).

³⁴¹ Heyman (2013, p. 1).

³⁴² Heyman (2009, p. 104).

³⁴³ Levy (2019, pp. 56–57).

³⁴⁴ Lewis (2015, p. 173). Numerous habit-based descriptions are found throughout Lewis’s book.

³⁴⁵ Wakefield (2017a, pp. 42–43).

susceptibility to use when cued/triggered; the risk of harmful consequences; the probabilistic nature (degrees) of control loss; the potential for controlled choice; and so on.

4.3.2 Capturing A Potluck of Views

Here I turn to a brief discussion of how the dispositionalist framework can accommodate other accounts of addiction. In other words, not only are most extant accounts already talking about a disposition, but the important components of those accounts can be incorporated into – that is, substantively unified under – the dispositionalist framework. This section explains how.

Recall that, on my view, an addiction disposition is a way some agent is, and being this way makes it true that the agent can (or *is disposed to*) behave in certain ways under certain conditions. Moreover, an addiction need not be realized even when in appropriate triggering conditions since it is plausibly a non-deterministic disposition. The physical makeup of the addict (a certain type of arrangement of the parts and qualities of the agent) will serve as the material basis grounding the addiction. There will also be paradigmatic realizations and manifestations of addiction that are triggered by being in certain types of circumstances. Additionally, as with many familiar skills or character traits, an addiction's development and maintenance are influenced by one's genes, upbringing, environment, exposure, pre-existing tendencies, and the like. As the blind men do with the elephant, addiction researchers tend to focus on one aspect of the whole phenomenon of addiction – development, environmental influences, psychological and behavioral manifestations, triggers, the brain (material basis), and so on. The following explains how each aspect fits into my framework.

First, consider the brain disease model that dominates the clinical literature. Proponents of this view focus on structural and functional changes to the brain, arguing that they become

dysfunctional in addiction.³⁴⁶ There are even non-disease or disease-neutral neurobiological models of addiction that aim to describe how addiction works at the level of the brain without committing one way or another to whether such workings are dysfunctional.³⁴⁷ On my view, these brain-based accounts are simply referring to (part of) the material basis of the addiction disposition. Of course, this does not make the neurobiology of addiction unimportant. We saw in the preceding chapter that dispositions are internally grounded, meaning that sufficient changes to the physical constitution of a person can generate, modify, or eliminate a disposition like addiction. This is precisely why pharmacological treatments can be effective in, for instance, helping to reduce addicted behaviors and even maintain abstinence by significantly minimizing cravings.³⁴⁸ However, anti-depressant medications that work on certain neurotransmitters and brain processes have also shown some success in reducing symptoms of depression, and yet this does not entail that all depression comes to is neurobiology. Neurobiological models of addiction help to discover which underlying physical changes to the brain are related to changes in whether, when, and how the disposition is developed, masked, manifested, modified, and so on. So, while we ought not identify addiction with brain states or processes, brain-based views are easily accommodated into the dispositionalist framework.

Second, consider views that focus on patterns of choice and behavior typical of addiction, such as those from George Ainslie and his followers.³⁴⁹ These authors focus on identifying and explaining addicts' habitual employment of certain choice principles and how this influences their subsequent patterns of behavior. Most then argue that such patterns of choice and behavior, while often harmful or irrational from a long-term perspective, are not abnormal and so cannot be indicative

³⁴⁶ Many examples exist in the literature (Hyman, 2005; Kalivas & Volkow, 2005; Kincaid & Sullivan, 2010; G. F. Koob et al., 2014; G. F. Koob & Le Moal, 2006; A. Leshner, 1999; Alan I. Leshner, 1997; N. Volkow, 2014; N. Volkow & Fowler, 2000; N. Volkow & Morales, 2015).

³⁴⁷ For instance, see Berridge & Robinson (2016).

³⁴⁸ See Froehlich et al. (2017), Jonas et al. (2014), Maisel et al. (2013), and Serecigni (2015).

³⁴⁹ See Ainslie (2017, 2019), Hanson (2009), and Heyman (2009, 2013, 2019).

of brain disease. Setting the merits of their arguments aside, on the account defended here these authors would simply be referring to *manifestations* of the addiction disposition, such as engaging in temporal discounting, failing to resist urges for imminent rewards, and experiencing harmful consequences. As I have pointed out, addiction is almost certainly a complex disposition that rests on further underlying dispositions, just as many skills or character traits do. Hence, it is worthwhile that researchers like Ainslie, Heyman, and other behavior-based rational choice theorists have identified some of these common underlying dispositions, such as a tendency towards exaggerated delay discounting. Nonetheless, we should see these views as merely focused on some of the ways in which the addiction disposition reveals itself, rather than claiming that the behaviors are the addiction.

Third, consider two sets of accounts that focus on various external and internal *influences* on addiction. On the one hand, biopsychosocial accounts influenced by the work of George Engel emphasize the extremely varied etiology (influences on the origination and development) of addictions.³⁵⁰ On the other hand, accounts like those from Hanna Pickard and Bruce Alexander that focus on addiction's context-sensitivity emphasize the importance of various contextual and environmental factors in influencing whether and how addiction is manifested and sustained, such as social pressure, poverty, opportunity, knowledge, denial, or self-identity.³⁵¹

The explanation of how the dispositionalist framework captures these views is again quite simple. Etiological accounts are explicitly intended to explain the *causes* of addiction, such as the role of genes, upbringing, comorbid diseases or disorders, and so on. Similarly, accounts focused on context-sensitivity are intended to explain the role of one's context, including features of both the external environment, such as opportunity, and of internal contingencies, such as how much one is in denial. Nothing about either type of account is unfriendly to a dispositionalist framework. Dispositions

³⁵⁰ See Buckner et al. (2013) and Marlatt et al. (1988).

³⁵¹ See Alexander (2010), Pickard (2019), and Sinnott-Armstrong & Pickard (2013).

originate from and develop according to causal influences, just like anything else. Moreover, dispositions are sensitive to context in a rather obvious way – certain types of circumstances *trigger* them to varying degrees. Hence, while these accounts are important in helping to map out some of the common developmental pathways and subsequent trajectories of an addiction, they are simply concerned with what my account would identify as causal influences or triggers of the addiction disposition.

It should now be clear that the dispositionalist account I am defending can play nicely with virtually any plausible view of addiction. This is because such accounts are simply focused on one or another aspect of the addiction disposition: material basis, manifestations, causal influences, triggering conditions, and so on. Addiction is a disposition of an agent. It is typically developed through various experiences and exposures that alter the physical makeup of the agent, such as undergoing characteristic changes to cognitive and motivational systems in the brain. This can be influenced by pre-existing tendencies, such as genetic makeup or distinct mental or behavioral disorders. Once present, we typically see the disposition manifested in certain patterns of choice and behavior in certain triggering circumstances. Moreover, which types of circumstances can trigger the disposition, and to what degree, can vary widely depending on the strength of the disposition, the arrangement of lower-level dispositions on which it rests, or the nature of the material basis that grounds it.

Before turning to a defense of premise (3d) in the next section, I will briefly demonstrate how the dispositionalist framework can help to expose merely verbal disagreements.

4.3.3 A Verbal Disagreement Exposed

Given the varied focuses of the views just discussed, it should now be clearer why it is so easy for proponents of different accounts to fail to make much dialectical progress. In short, many authors often talk past one another. This is somewhat unsurprising given that each view places emphasis on

very different kinds of entities in reality. In this section, I focus on a particular disagreement between Nora Volkow and Gene Heyman over the disease status of addiction.

Nora Volkow argues that addiction is a brain disease.³⁵² She appeals to numerous well-established neurobiological studies that have generated large sets of data about the functioning – or, for Volkow, *dys*functioning – of different mechanisms and processes in the brains of addicts. For Volkow, the neurobiological studies show that addicts’ brains are functioning abnormally, and she ultimately concludes that the abnormalities are pathological. Hence, addiction is a brain disease.

The Harvard psychologist Gene Heyman argues that addiction is not a brain disease.³⁵³ He appeals to various well-established epidemiological and behavioral studies that have generated large sets of data about recovery rates and choice patterns (and thus behavioral patterns) of addicts. For Heyman, the studies he appeals to show that addicts usually recover without treatment, for normal reasons, and that they can exercise what appears to be a capacity for voluntary, controlled choice. Moreover, he argues, the choice patterns we see in addiction can be normalized since they can be predicted by choice models from behavioral economics that explain most of our ordinary, non-addicted behaviors.³⁵⁴ Hence, addiction is not a brain disease.

So, what is the problem here? Even setting aside the question of whether Volkow and Heyman understand disease in the same way (they do not), a glaring problem of talking past one another remains. The dispositionalist account helps to bring out where Volkow and Heyman are missing each other, and why their views are not as incompatible as they might initially seem to be. When Volkow says ‘addiction is a brain disease’, this can roughly be translated as ‘the characteristic functional and structural changes in the brains of addicts constitute a disease’. When Heyman says ‘addiction is not a

³⁵² See Volkow’s various works (R. Z. Goldstein & Volkow, 2002; Kalivas & Volkow, 2005; N. Volkow, 2014; N. Volkow et al., 2016; N. Volkow & Fowler, 2000; N. Volkow & Koob, 2015; N. Volkow & Morales, 2015).

³⁵³ See Heyman (2009, 2013, 2019).

³⁵⁴ Two examples are matching law (https://en.wikipedia.org/wiki/Matching_law) and melioration theory (https://en.wikipedia.org/wiki/Melioration_theory).

brain disease’, this can roughly be translated as ‘the characteristic behavior and choice patterns of addicts do not constitute a disease’. Volkow is referring to (part of) the material basis of the addiction disposition. Heyman is referring to the realizations or manifestations of the addiction disposition. Contrast Heyman’s position with Marc Lewis’s response to Volkow.³⁵⁵ A neuroscientist himself, Lewis accepts all of the neurobiological data and yet still argues that addiction is not a brain disease. However, he does so by arguing that the functional and structural brain changes characteristic of addiction are *not* dysfunctional. Thus, he and Volkow are speaking about the same phenomena – the material basis of the disposition – and so have a genuine disagreement over its status as dysfunctional.

Heyman’s focus on the behavioral manifestations of addiction is not irrelevant; these are important matters. The problem is that he and Volkow are simply not talking about the same thing, and so their claims about addiction and disease end up sailing past one another like ships in the night. Understanding the various components of the dispositionalist framework helps us to tease apart what exactly these authors are getting at in their accounts. It is in this way that understanding addiction as a disposition can help to expose and ultimately avoid merely verbal disagreements, and can reveal how two apparently competing accounts might be reconciled under a single framework.

4.3.4 A Unifying Account

The aim of Section 4.3 has been to show how the dispositionalist account can unify some prominent views in the literature. Though I lack the space here for a fuller exploration, I believe such unification can be extended to include many more accounts. Although it is typically implicit, many accounts already speak in dispositionalist terms. Moreover, where they do not, the core of the accounts can still be brought into the dispositionalist framework. When this fact about the literature is coupled with the fact that disposition-friendly language and features of the account are quite often implicit and

³⁵⁵ See Lewis (2015).

unnoticed, it makes the development of a dispositionalist framework all the more important. Even more explicitly dispositionalist accounts like those from Sinnott-Armstrong and Pickard and Foddy and Savulescu still fail to provide any worked-out account of the dispositionalist components of their views. This is the gap I am aiming to fill, and in doing so I have shown how such an account can provide substantive unification to most views on offer.

4.4 Ontologizing the Dispositionalist Account

The previous chapter defended premise (3a), which states that there is a further, distinct sense of ‘unification’ that ontology alone cannot provide: *substantive unification*. Through the defense of premises (3b) and (3c) in the preceding two sections, we can now more clearly see why the dispositionalist account can be said to provide this. Specifically, it helps to unify the literature over and above the methodological unification provided by the implementation of ontology.

Now, given my defense of the importance of implementing ontology into the addiction literature, this section defends premise (3d), showing how my account fits with the methods of ontology building – specifically, ontologies conformant with Basic Formal Ontology (BFO). First, drawing on the principles of best practice discussed in Chapter 2, I explain how the dispositionalist account of addiction defended here is in line with proper ontology building. In this way, the section demonstrates how my account meets important desiderata that facilitate the methodological unification provided by ontology. If it did not, then premise (3a) would be false since the dispositionalist account would not provide *further* unification to the literature were it inconsistent with the practice (ontology) that provides the initial methodological unification. Second, I close the section with an example of how my account would start to be represented (that is, diagrammed) ontologically.

4.4.1 Adhering to Ontology Best Practices

The principles of best practice underlying BFO are intended to facilitate methodological unification. Hence, given my argument that ontology should be applied to the addiction literature, it is important to demonstrate that my account indeed follows these principles. I lack the space to discuss each principle, and hence I only focus on some of the more central ones covering the following topics:

- (i) constructing definitions;
- (ii) the backbone *is_a* hierarchy; and,
- (iii) realism and linking the ontology back to BFO.

4.4.1.1 Proper Definitional Form

Definitions should be in Aristotelian form (**A** *is_a* **B** *that Cs*) and, insofar as possible, appeal to terms that are either simpler or can themselves be clearly defined in simpler terms, while avoiding circularity. The definition of ‘addiction’ I defend meets all of these criteria: **ADDICTION** *is_a* (systematic) **DISPOSITION** *that is realized in failures to control one’s desires to engage in certain types of behavior*. Systematic dispositions are currently unaccounted for in BFO or any other ontology. However, I have provided an account of systematicity. These are dispositions that are sufficiently strong (significantly likely to be realized when triggered) in a sufficient portion of those triggering conditions that members of the bearer’s reference class have a sufficiently high probability of being in, given the history and laws of nature. One might worry about the threshold issues raised by this definition. However, it is important to note that this is distinct from the definition appealing to concepts that are unclear. We know what this definition means, even if we do not yet know where the cutoff is. The definition also avoids circularity.

Regarding desires and control, currently only a few ontologies in the OBO Foundry attempt to represent these entities. Here are the definitions and *is_a* classifications from these ontologies:

Semantiscience Integrated Ontology (SIO):³⁵⁶

- (i) **DESIRE** =def. *A strong emotion of wanting to have something or wishing for something to happen.*
 - a) *is_a* **INTEREST** (SIO), which *is_a* **POSITIVE EMOTION** (SIO), which *is_a* **EMOTION** (SIO), which *is_a* **BEHAVIOUR** (SIO), which *is_a* **PROCESS** (SIO)

Experimental Factor Ontology (EFO):³⁵⁷

- (i) **CONTROL** =def. *The act of directing or determining; regulation or maintenance of a function or action; a relation of constraint of one entity (thing or person or group) by another. A control role is borne by a material in a process in which results obtained from an experimental sample and a control sample are compared.*
 - a) *is_a* **CHEMICAL ROLE** (EFO), which *is_a* **ROLE** (EFO), which *is_a* **MATERIAL PROPERTY** (EFO)

The Gender, Sex, and Sexual Orientation Ontology (GSSO):³⁵⁸

- (i) **CONTROL** =def. *modification of the execution of an event or process.*
 - a) *is_a* **RELATIONSHIP** (GSSO), which *is_a* **CHARACTERISTIC** (GSSO), which *is_a* **QUALITY** (BFO)
- (ii) **DESIRE** =def. *A strong emotion of wanting to have something or wishing for something to happen.*
 - b) *is_a* **INTEREST** (SIO), which *is_a* **POSITIVE EMOTION** (SIO), which *is_a* **EMOTION** (SIO), which *is_a* **MENTAL PROCESS** (NCIt³⁵⁹), which *is_a* **NERVOUS SYSTEM PROCESS** (GO), which works its way up the *is_a* hierarchy through other Gene Ontology terms to **BIOLOGICAL PROCESS** (GO), which *is_a* **PROCESS** (BFO)

There are a number of problems with the way control and desire are represented in these ontologies. However, of importance is that there is at least some content here that can be worked with to clarify these terms so that the definition of ‘addiction’ appeals to concepts we can understand.

For instance, the GSSO classifications of **CONTROL** and some of its sibling classes suggest that control might be what is called a ‘process profile’ in BFO. Roughly, a process profile is analogous to a property of a process, such as its *speed* or *rate*.³⁶⁰ Similarly, perhaps control is something like a *way*

³⁵⁶ See classes for SIO here: <https://biportal.bioontology.org/ontologies/SIO>. For the hierarchical classifications identified in the lists here, I put the acronym of the ontology that hosts the parent class in parentheses. This is to help show when the ontology does (and does not) link back up to BFO, for instance.

³⁵⁷ See classes for EFO here: <https://biportal.bioontology.org/ontologies/EFO>.

³⁵⁸ See classes for GSSO here: <https://biportal.bioontology.org/ontologies/GSSO>.

³⁵⁹ This is the National Cancer Institute Thesaurus, which is a taxonomy of terms that started as a thesaurus from the NCI and is trying to become an OBO Foundry ontology. I return to NCIt in the ontology evaluations in Chapter 5.

³⁶⁰ Technically, processes cannot bear properties in BFO; only the material entities that participate in them can. However, there does seem to be a sort of analogy, such as the process of my running around the block unfolding at a particular rate. This is precisely why *process profile* was introduced. See Smith (2012) for discussion of process profiles.

in which a process unfolds, loosely characterized as a sort of property certain processes like actions might have. So, just as one way that my running might unfold is such as to be *medium paced*, a way my choice might unfold is such as to be *uncontrolled* (or perhaps *medium controlled*). Another possibility is that control amounts to one process regulating another, where this is understood as the controlled process being caused or constrained by the right kind of (controlling) process. Both EFO and GSSO gesture at regulation as an understanding of control. This idea could line up with causal theories of agency and control in the philosophy of action literature.³⁶¹

Both SIO's and GSSO's representations of desire are also similar to the reward theory of desire from Arpaly and Schroeder discussed in the previous chapter. Each sees desire as a process – a *desiring*. Arpaly and Schroeder see *desiring* as the process of certain mental representations playing a causal role in the calculation of an overall reward value, ultimately leading to the modification of the agent's dispositions. This seems especially amenable to the GSSO classification, which invokes biological and mental process terms from GO referencing the brain and nervous system. The upshot is that we have the ingredients to flesh out desire and make the definition of 'addiction' understandable.

4.4.1.2 A Consistent “is a” Hierarchy

My account conforms to the requirements that ontologies must have a consistent *is_a* hierarchy as its backbone. Put very simply, **ADDICTION** is classified as a *subtype_of* BFO **DISPOSITION**, thereby linking it directly up to BFO. Moreover, this means that **ADDICTION** inherits all of the essential properties of **DISPOSITION** according to BFO. This includes having a material basis, certain types of processes that realize it, certain types of circumstances in which the realization can occur, and so on. Expanding this further, different kinds of addictions, such as various substance and behavioral

³⁶¹ Donald Davidson's classic account is one example (Davidson, 2001). John Martin Fischer's (1982) *reasons-responsiveness* account and Carolina Sartorio's (2016) *actual causal sequence* account of control are two examples in the free will literature.

addictions, would count as subtypes of **ADDICTION**. Following the transitivity of the *is_a* relation, these will also be subtypes of **DISPOSITION**.

This seems quite simple, but it is an important point to get right ontologically.³⁶² Consider, for instance, that brain disease models of addiction have been accused of having trouble accounting for certain behavioral addictions. This is because the latter lack the ingestion of a chemical compound that works directly on the reward system in the brain, which is an important part of the brain disease model's explanation of how addiction works and, most importantly, what makes it a disease.

Finally, my account separates triggering conditions, manifestations, causes, and other influences on addiction from the addiction disposition itself. In doing so, it avoids the problem of classifying processes as dispositions, material entities as processes, and other kinds of category mistakes that existing accounts tend to make.

4.4.1.3 Realism and BFO

As mentioned, my account of addiction allows it to link straightforwardly to BFO because, at the most fundamental level, I take addiction to be a disposition. This is important because BFO (and various BFO-conformant ontologies like GO) set the standard for good ontology building. Hence, defining 'addiction' in a way that is so fundamentally aligned with BFO ensures that my account will have very little trouble remaining BFO-conformant.

One of the most important ways in which an ontology ought to be conformant with BFO is to adhere to the principle of realism. That is, terms should refer to entities and relations in the world (as opposed to concepts or ideas). It should be fairly clear that the account of addiction I defend is indeed realist. On my view, addiction is a disposition of an agent (typically an organism), where this is

³⁶² Indeed, the transitivity of the *is_a* relation happens to be a principle that many extant ontologies fail to respect. When ontologies are poorly built and end up dying, a failure to maintain a consistent *is_a* hierarchy is typically one of the main reasons. Being unaware of or unconcerned about *is_a* transitivity is often a big part of this.

a *way* that agent is. Moreover, the disposition is grounded in the actual physical makeup of the agent, the circumstances that trigger the disposition are states and processes in the actual world, and the realization and manifestations of the disposition are actual processes and states in the world. The dispositionalist account of addiction defended here is realist through and through.

Now that we have seen how the dispositionalist account is in line with the core ontology principles of best practice, I turn to how the account would be represented in an ontology.

4.4.2 Representing the Components of the Dispositional Account in a BFO-Compliant Way

An ontology is a kind of model or *representational artifact*. Hence, if an account is to fit with the methods of ontology, then we ought to have a sense of what it would look like to model it ontologically. One thing this can help us to do is to see (somewhat literally) where and how other accounts would fit into the model. Visualizations can give us a better sense of what (or where) in the world addiction is. The node-and-edge graph models of ontology do this by visually representing which types of entities are involved (nodes) and how they are interrelated (edges).

Before we start to represent addiction, two caveats are in order. First, the diagrams presented here are not meant to be the final say on ontologizing addiction. They are, at most, revisable suggestions for how the relevant entity types might be classified and interrelated in an ontology. In this way, I respect the *open world assumption* underlying BFO, roughly stating that our ontological assertions are fallible and so always open to revision. This is especially so in early stages of the process. Second, we will not see every aspect of addiction modeled here. Dispositions are relatively complex entities ontologically, being intimately related to various processes, material entities and their parts, environmental features, and so on. What is more, addiction in particular is quite a complex disposition. Hence, there are many possible diagrams that might capture the intricacies of the phenomena surrounding the addiction disposition. I do not have the space to explore them all, and so I focus only

on some of the core features of the account. Representing and refining the many aspects of the account is an ongoing part of the project undertaken here. Such is the open-ended nature of ontology.

4.4.2.1 Addiction as a Disposition in BFO

At the most fundamental level, **ADDICTION** *is_a* **DISPOSITION**. Dispositions are well-established entities in BFO, and so representing this is quite simple. This is seen in Figure 12 below.³⁶³

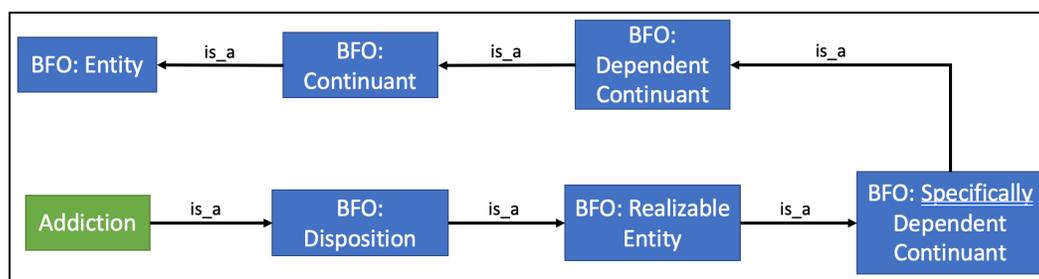


Figure 12: *Addiction as a Disposition in BFO*

As we know by now, that some entity is a disposition entails a number of things about it, such as that it is realized in processes of a certain type, the physical makeup (parts and qualities) of its bearer serves as its material basis, and so on. Let us see how these various other components of the dispositionalist account might be represented.

4.4.2.2 Addicts Are Organisms Bearing the Addiction Disposition

Addiction will always have some bearer, and that bearer is the *addict*. Plausibly, this will be an organism since it seems like anything that has desires, engages in behaviors, attempts to control that behavior, and so on is very likely an organism of some kind.³⁶⁴ The Ontology of Biomedical Investigations (OBI) includes **ORGANISM**, defined as “a material entity that is an individual living

³⁶³ For the diagrams in this section, I use green boxes to indicate that a term is either not yet in an established, BFO-conformant ontology or, if it is, it is still in need of being worked out. I use blue boxes to indicate which ontology a term is from using the ontology’s acronym. All relations (edges) are labeled with the type of relation being represented. Insofar as is possible, I will use only those relations which are already defined in BFO or the Relation Ontology (RO).

³⁶⁴ We can also think of the bearer more generally, such as an agent. I will stick with organism for the diagramming here since it is most common and there is a corresponding class in BFO-conformant biological ontologies.

system, such as animal, plant, bacteria or virus, that is capable of replicating or reproducing, growth and maintenance in the right environment.”³⁶⁵ This is represented in Figure 13 below.

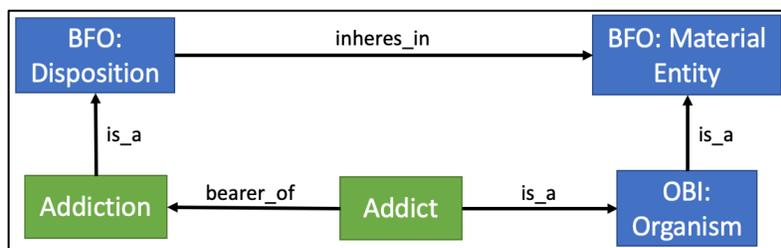


Figure 13: Addict as the Bearer of the Addiction Disposition

If it is possible than non-living entities can be addicted, such as artificial intelligence systems, then the parent class of **ADDICT** could simply be **MATERIAL ENTITY**. Alternatively, the Common Core’s Agent Ontology has the class **AGENT**, roughly understood as a **MATERIAL ENTITY** *that is capable of performing intentional acts*. This might be able to accommodate non-living, goal-oriented entities like artificial intelligence systems.³⁶⁶ In this case, **ADDICT** could simply be a *subtype_of* **AGENT**.

4.4.2.3 Addiction Is Realized in Uncontrolled Behaviors

Addiction is realized in the process of failing to control one’s desires. Of course, we also know that not just any instance of failing to control one’s desires counts as the realization of an addiction. It is only those failures that are the realization of the *systematic disposition* to fail to control one’s desires. To capture this, we can call any failure to control one’s desires that is in fact a realization of an addiction an ‘addicted behavior’. Each will still be a *subtype_of* of **BEHAVIOR**, which is in the Gene Ontology Extension (GOE) and is defined as, “The specific actions or reactions of an organism in response to external or internal stimuli. Patterned activity of a whole organism in a manner dependent upon some combination of that organism’s internal state and external conditions.”³⁶⁷ In short, an

³⁶⁵ See the definition and OBI’s other classes here: <https://bioportal.bioontology.org/ontologies/OBI>.

³⁶⁶ I am unaware of a website or published paper that provides the definitions for terms in the various Common Core Ontologies. However, an excel file with the terms and their corresponding definitions and parent classes can be made available upon request if the reader is interested.

³⁶⁷ See the definition of ‘behavior’ and other GOE classes here: <https://bioportal.bioontology.org/ontologies/GO-EXT>.

addiction is *realized_in* a certain kind of behavior, which will be some failure to control one’s desires. The addict bearing the addiction will be a *participant_in* the behavior. Figure 14 below represents this.

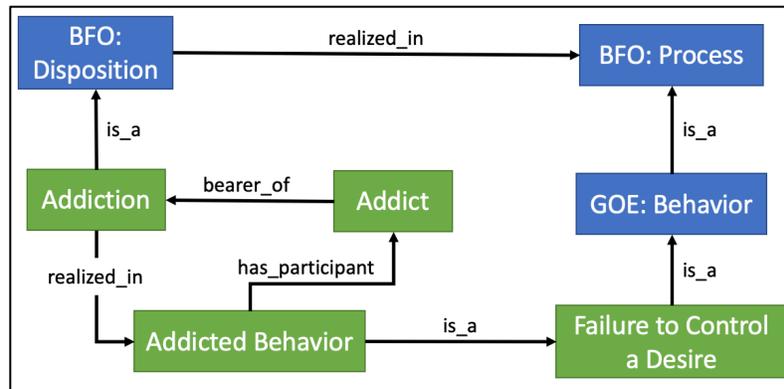


Figure 14: Addiction as Realized in Addicted Behavior

One might worry that the class **FAILURE TO CONTROL A DESIRE** is inconsistent with the realism underlying BFO since ‘failure’ makes it seem to be about a process that never happens – that is, a non-existent entity. However, appearances can be deceiving. It is not some non-existing entity which this class picks out. Instead, the failure of control is a behavior performed by an agent, but which lacks a certain feature that is common of many ordinary behaviors, namely, control. In this way, it is not a complementary class like **NONDOG** that asserts a negative universal. Instead, it uses negative information to define a class representing existing entities (like uncontrolled behaviors or missed putts). Smith and colleagues admit just this kind of exception for using negative terms in an ontology, as in their example of a prokaryotic cell:

There are, however, some cases where classes involving a negative element in their definition will properly be included in an ontology. Thus, for example, prokaryotic cells are distinguished from eukaryotic and all other cells precisely by the fact that they lack a cell nucleus. This is, in effect, negative information used to define a class...It is just that the definition of these cells itself includes some negative information (that they are cells that do not have a nucleus)...The recommendation to avoid negative terms thus needs to be applied with care...³⁶⁸

We could have an analogous understanding of, for instance, the class **SUBSTANDARD PUTTING ABILITY**, which we could define as being *realized_in* failures to make putts. We do not need to invoke

³⁶⁸ Arp, Smith, & Spear (2015, p. 75).

non-existing entities or absences here since the failed putts are failures to make *attempted* putts. Thus, rather than an absence, a process like **POORLY EXECUTED PUTT** would realize this disposition.

4.4.2.4 Attempts to Control One's Desires in Societally Relevant Choice Situations

Consider next that dispositions like addiction also have certain triggering conditions. A golfer's putting expertise is triggered by attempting to make putts in certain conditions (and then realized, with some sufficiently high probability, in sinking them). Similarly, an addict's addiction is triggered by attempts to control their desires in certain conditions (and then realized, with some sufficiently high probability, in failures to do so – uncontrolled behaviors). On the dispositionalist account I defend, these certain conditions are societally relevant choice situations. That is, they are situations in which an opportunity arises to make a choice that is relevant to the object of addiction, such as ingesting alcohol, engaging in gambling, throwing away one's cigarettes, signing up for rehab, and so on. Moreover, these choice situations are *societally relevant* – that is, members of the addict's reference class have a sufficiently high probability of being in them, given the laws and history of the world.

There are three main components here. The first two concern the triggers and the triggering conditions of addiction. Attempts to control one's desires are behaviors. As their name indicates, choice *situations* are a kind of context, situation, setting, or environment. We can appeal to the class **CONTEXT** from the Cognitive Atlas Ontology (COGAT), which is defined as “a set of interrelated conditions in which something exists or occurs.”³⁶⁹ While the definition is not perfect, it is a useful start. Similarly, in defining ‘physical setting’ for an intervention, the developers of the Behavior Change Intervention Ontology begin with the definition of ‘setting’ as “the place or type of surroundings where something is positioned or where an event takes place.”³⁷⁰ In both cases, the choice situation that makes up part of the triggering conditions would be a kind of **OBJECT AGGREGATE**, located in

³⁶⁹ See the definition and COGAT's other classes here: <https://biportal.bioontology.org/ontologies/COGAT>.

³⁷⁰ Norris et al. (2020).

some spatial region, and typically in which some process occurs. The trigger and triggering conditions are represent in Figure 15 below.

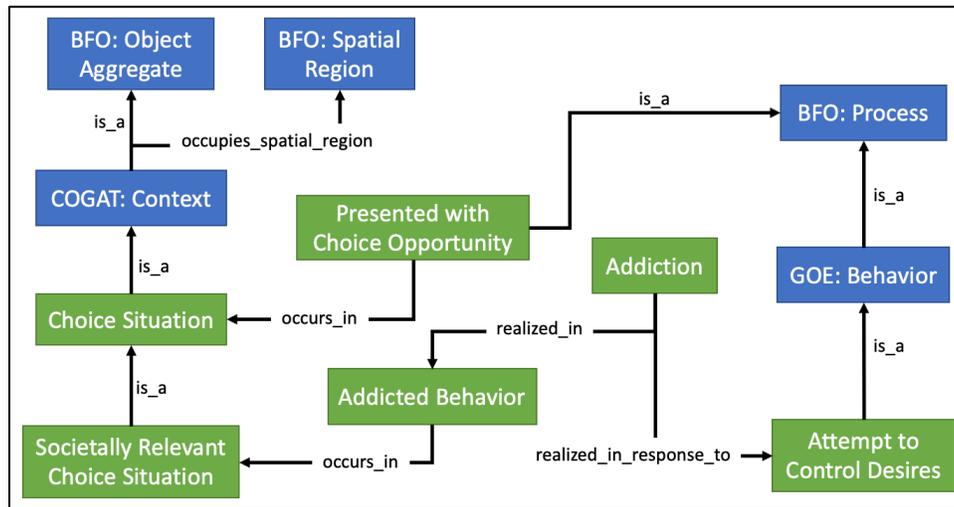


Figure 15: Addiction as Triggered by Attempts to Control Desires in Choice Situations

The third component is the idea that the choice situations are *societally relevant*. To represent this, we need to include a reference class for the addict, which would be a group of organisms (or perhaps a group of agents). This would be a kind of **OBJECT AGGREGATE** in BFO. The Psychology Ontology (APAONTO) contains the class **REFERENCE GROUP** that could serve as a good start. It is defined as, “social groups used as sources for personal and behavioral identification, motivation, and evaluation of one’s own status.”³⁷¹ Furthermore, we need to represent the fact that these choice situations are such that members of the addict’s reference class have a sufficiently high probability of being in them. There is currently no existing relation that easily expresses this, and so I have simply asserted a relation that captures the idea. Ultimately, this may require having statistical data on the individual instances that are members of an addict’s actual reference class. However, I only offer an initial attempt at representing this to begin the development process that will eventually be needed. Figure 16 below represents the entities and relations surrounding societally relevant choice situations.

³⁷¹ See the definition and APAONTO’s other classes here: <https://bioportal.bioontology.org/ontologies/APAONTO>.

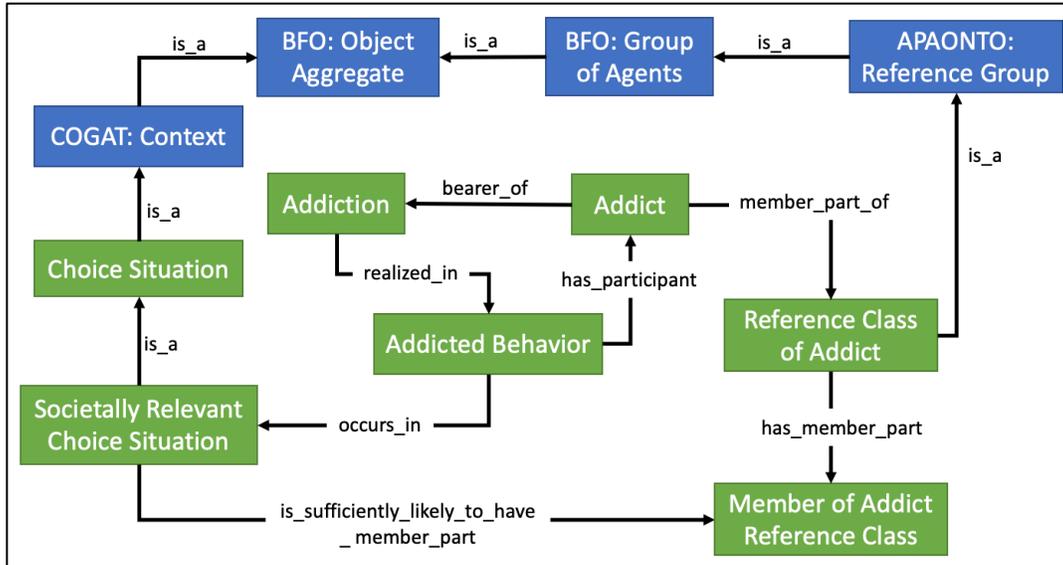


Figure 16: Societally Relevant Choice Situations as Related to the Addict's Reference Class

These are the beginnings of representing addiction in a BFO-compliant ontology. As mentioned above, ontology building is a collaborative, open-ended developmental process. As such, nothing here is set in stone. The point is to illustrate both *that* my account can work with the methods of ontology, and *how* it can begin to do that. Let us turn briefly to discussing why the dispositionalist account is not only *workable* within the methods of ontology building but also *useful*.

4.4.3 The Ontological Usefulness of the Dispositionalist Account

I argued above that the dispositionalist account helps to unify extant views of addiction. This already makes the account quite useful (for instance, for purposes of communication and conceptual clarity). Still, this section argues that the dispositionalist account (and its unifying power explained above) is also useful in a way that more straightforwardly applies to the use of ontologies.

One of the main purposes (if not *the* main purpose) of an ontology is using it to tag the data and literature relevant to the domain the ontology is about. The Gene Ontology is used to tag biological and genomic data and literature, the Ontology for General Medical Science is used to tag medical data and literature, and so on. Data and literature tagging are incredibly important – and so part of what grounds the importance of ontology – because this allows for that data and literature to

be more effectively and efficiently discovered, queried, merged, and analyzed. Perhaps most importantly of all, it allows this to happen *even when the data and literature are vast and heterogenous*. Hence, ontologies are critical in the scientific endeavor to understand reality as deeply, clearly, and comprehensively as we can. So, why does this matter to the dispositionalist account of addiction?

As I have pointed out, an important feature of ontology is that it is neutral with respect to the *content* of different views. Robert West, Janna Hastings, and their colleagues noted that this is a positive feature since it means that multiple, competing perspectives on addiction can all be captured. If one author claims addiction is a behavior, such as repeated failures to quit using despite harmful consequences, while another claims it is a disease in the brain, this is unproblematic for an ontology. Each account would simply define ‘addiction’ in terms of their respective views, and we would see that one is referring to an aggregate of processes (a behavioral pattern) and the other is referring to a disposition inhering in a particular anatomical structure (a disease in the brain). Moreover, we could give each a distinct label (such as ‘behavioral pattern addiction’ and ‘brain disease addiction’) with, most importantly, a unique identifier (a URI). Thus, we can easily disambiguate the data and literature corresponding to each view, we do not force different perspectives to align, and we respect the principles of *perspectivalism* and *adequatism* underlying BFO.³⁷²

However, the problem is that this does not account for the fact that most views are talking about the same thing when they speak of addiction. To be sure, many authors focus on different *aspects* of addiction, often according to their discipline or training. However, the dispositionalist account helps us to see that these are almost always aspects of the same thing – the addiction disposition. To use the above example, viewing addiction as a behavioral pattern is to focus on the various *behavioral manifestations* of addiction, while the brain disease view is focused on part of the *material basis* of

³⁷² As a reminder, *perspectivalism* states that there are multiple accurate descriptions of reality and *adequatism* states that the entities in a given domain should be taken seriously on their own terms, and not viewed as reducible to other entities.

addiction. As another example, the proponent of the biopsychosocial model is focused on the various *causal influences* on the development or manifestations of an addiction. There is good reason to believe that seemingly disparate accounts of addiction are, with respect to *what addiction is*, not so disparate. So, it is preferable to avoid having multiple definitions of ‘addiction’ corresponding to each disciplinary perspective. This does not thwart *perspectivalism* or *adequatism*. The various aspects, at their various levels of granularity (neuron, desire, behavior, context), can all stand on equal footing ontologically. Instead, it is simply to recognize that these are indeed merely *aspects* or *components* of addiction, and not addiction itself. The dispositionalist account I defend, unlike ontology by itself, helps us to see that.

The upshot, then, is that the data and literature relevant to addiction need not be bombarded with as many definitions of ‘addiction’ as there are various perspectives and approaches to studying it. We can tag the data and literature with the various terms (and definitions) corresponding to the many components involved in addiction. However, the dispositionalist account allows us to remain precise about what in the world we identify as addiction. Again, *if* someone identified addiction with, for instance, brain states, then the ontology could still capture this. However, I have argued that we have better reason to believe that most researchers already are (when charitably interpreted) or would be (if pushed) very much on the same page about the fundamental, dispositional nature of addiction.

4.5 Defending the Bridge Premise (3e) of the Argument for Premise (3)

Let us take stock of the defense of premise (3) in my overarching argument. Here again is the argument for premise (3):

- (3a) Ontology cannot *fully* unify the literature (that is, there is a further, distinct sense of ‘unification’ that ontology alone cannot provide); [Chapter 3]
- (3b) The dispositionalist account of addiction best explains the key phenomena surrounding addiction; [Section 4.2]
- (3c) The dispositionalist account of addiction captures other competing theories and can explain their (often apparent) disagreements; [Section 4.3]
- (3d) The dispositionalist account works with the methods of ontology; [Section 4.4]

- (3e) If (3a)-(3d), then the dispositionalist account of addiction is true and provides still further unification to the literature; [Section 4.5]
- (3) Hence, the dispositionalist account of addiction is true and provides still further unification to the literature. [(3a)-(3d)]

Premises (3a)-(3d) have been defended. The present section defends premise (3e), which acts as the bridge between the other premises and the conclusion (3).

The conclusion makes two claims. First, that the dispositionalist account of addiction is true. This is supported by premises (3b), that this account best explains key phenomena, and (3c), that it also captures other accounts and can help to explain their disagreements. The move from (3b) to the first part of (3) is simply an inference to the best explanation. The best explanation of some phenomenon (or phenomena) is the actual explanation, and the actual explanation is necessarily true since it accurately describes reality.³⁷³ I argued above that understanding addiction as a disposition – and, in particular, a *systematic disposition to fail to control one's desires* – best explains what we know about addiction. It is consistent with the empirical data, it avoids counterexamples, and so on. Hence, this is reason to believe the dispositionalist account *actually* explains the relevant phenomena, and therefore, that what the account postulates, such as that addiction is a disposition, is *true*.

The move from (3c) to the first part of (3) also invokes an inference to the best explanation. To capture other extant accounts was, in a way, to explain what is true about them. More than this, the explanatory power of the account included its ability to show that, on closer inspection, seemingly competing accounts are often talking about the same thing (a disposition) and thus are not really in competition at all. This is why an upshot of (3c) was that the dispositionalist account brings *substantive unification* to the literature. But just as the truth of an account is supported by how well it explains the relevant phenomena, it seems to me that its truth is also supported by how well it captures what

³⁷³ For discussion of inference to the best explanation and its connection to truth, see Cline (2016, Sect. 4).

existing accounts get right and avoids what they get wrong. Thus, (3c) also supports the first part of the conclusion (3).

The second part of the conclusion states that the dispositionalist account of addiction provides unification to the literature over and above that provided by implementing the methods of ontology. This was supported by the joint work of premises (3a), (3c), and (3d). As just explained, premise (3c) was about the dispositionalist account providing substantive unification to the literature. The truth of premise (3a) entails that this unification is distinct from the methodological unification ontology provides, and that ontology cannot *also* provide substantive unification. Finally, premise (3d) secured the fact that the dispositionalist account defended here is indeed workable within the method of ontology and its principles of best practice. Since this account provides a kind of unification that ontology cannot, and since providing that unification is not inconsistent with implementing ontology and securing the methodological unification it provides, it follows that the second half of the conclusion is true. That is, the dispositionalist account provides *still further* unification to the literature.

Putting all of this together, the truth of premises (3a)-(3d) support the conclusion that is (3). Hence, the bridge premise (3e) is true, as is premise (3) of my overarching argument.

4.6 Lessons and Looking Ahead

This chapter has explained and defended the power of the dispositionalist account of addiction. However, rather than a kind of despotic and tyrannical power, my account aspires to a more democratic rule. I claimed that, while it may sound surprising, most accounts of addiction seem to be getting at the same kind of thing in reality when we focus on what they take addiction to be. Sometimes this is hard to see since, whether intentionally or not, addiction researchers can lose sight of the *nature* (or metaphysics, or ontology) of addiction itself for the *components* of addiction, such as how it works, what influences it, and the like. Understanding addiction as a disposition lets us see that most extant

accounts can still have a voice under one and the same framework – hence the *democratic* nature of its power. That said, I also tried to show that the dispositionalist account should still rule, even if only democratically. Extant accounts can be unified precisely because there is an ontological common ground residing underneath, and that common ground is the disposition that is addiction. Thus, while the metaphor of the blind men and the elephant is apt in certain ways for capturing the state of the literature, we need not make relativistic or even pluralistic conclusions on its basis.

In addition to exhibiting the power of accepting that addiction is a certain type of disposition involving impaired control over one's desires, I argued that this view is compatible with the methods of ontology. More than this, though, I argued that the account is actually useful ontologically speaking, and I provided an initial glimpse of what it would look like when modeled in a BFO-conformant ontology. This part of the project was not intended to be complete since ontology is an ongoing process. However, it is useful to see where the core components presently stand ontologically.

In the next chapter, I turn to defending premise (4) of my overarching argument, which states that the dispositionalist account I have been defending fares better than existing accounts of addiction. The present chapter included a few comparisons with other authors' views of addiction. However, Chapter 5 does two additional things. First, it focuses more precisely and methodically on identifying particular problems with a handful of prominent existing accounts, and why the account I defend avoids these problems. While we saw above how the dispositionalist account can capture relevant phenomena (such as the puzzle of addiction), the next chapter focuses on particular mistakes within popular accounts. Second, the next chapter adds an evaluation of representations of addiction in existing ontologies. In the end, we will see that the dispositionalist account indeed fares better than the others on offer.

Chapter 5: How the Dispositionalist Account Fares Against Extant Accounts

5.1 Introduction and Chapter Road Map

Up to this point, I have motivated the need for my project, explained and defended the role of ontology in that project, and explained and defended the dispositionalist account of addiction, including its unifying power. It is now time for the comparative component of my project. First, as a reminder, the overarching argument I have been defending is restated below:

- (1) The addiction literature suffers from a serious problem of disunification; [Chapter 1]
- (2) Ontology can help to solve the addiction literature's problem of disunification; [Chapter 2]
- (3) The dispositionalist account of addiction is true and provides still further unification to the literature; [Chapters 3 and 4]
- (4) The dispositionalist account of addiction fares better than competitor accounts, both philosophical and ontological; [Chapter 5]
- (5) If (1)-(4), then addiction research ought to develop and implement an ontology of addiction and adopt a dispositionalist account of addiction; [trivial]
- (6) Hence, addiction research ought to develop and implement an ontology of addiction and adopt a dispositionalist account of addiction. [(1)-(5)]

The preceding chapters defended the first three premises of this argument. This chapter defends the fourth by evaluating extant views in the literature, and then explains why the fifth allows an easy move from the premises to the conclusion. It is structured as follows.

First, Section 5.2 presents two sets of criteria on which my evaluations will be based. The first includes what I call 'substantive desiderata' (or 'substantive criteria') since they pertain to what a satisfactory account of addiction ought to look like. Some of this will be familiar from Chapters 1 and 4. The second includes what I call 'methodological desiderata' (or 'methodological criteria') since they pertain to following the principles of best practice for BFO-conformant ontology building. That is, the second set of criteria is about what a satisfactory ontological representation of addiction ought to look like. Much of this should be familiar from Chapter 2.

In Section 5.3, I survey a number of prominent views of addiction in the (non-ontology) literature. I show how each is flawed, failing on one or more of the just-mentioned substantive criteria, and I note how the dispositionalist account avoids these flaws. This section complements the discussion in the preceding chapter that explained how the dispositionalist account best explains key phenomena surrounding addiction. That discussion focused more on the positive features of the dispositionalist account (where this sometimes meant capturing what extant accounts get right). Accordingly, Section 5.3 focuses on the negative features of extant accounts, bolstering the claim that the dispositionalist account fares better than these extant accounts of addiction.

Section 5.4 is similar in structure, though the focus there is on evaluating existing *ontological representations* of addiction.³⁷⁴ I focus on a number of ontologies managed by the Open Biological and Biomedical (OBO) Foundry that attempt to include the class **ADDICTION** (or some class for which ‘addiction’ is an intended synonym, such as **SUBSTANCE USE DISORDER**). As above, I show how each is flawed by failing on one or more of the above-mentioned methodological criteria. Moreover, since the definitions of terms in an ontology are in some sense an account of the entities they represent, I also show how these ontological representations fail on one or more of the above-mentioned substantive criteria. Thus, the argument for premise (4) will look like this:

- (4a) The dispositionalist account fares better than prominent extant competitor views with respect to satisfying the substantive desiderata; [Section 5.3]
- (4b) The dispositionalist account fares better than existing ontological representations of addiction with respect to satisfying the substantive and methodological desiderata; [Section 5.4]
- (4c) If (4a) and (4b), then the dispositionalist account of addiction fares better than competitor accounts, both philosophical and ontological; [trivial]
- (4) Hence, the dispositionalist account of addiction fares better than competitor accounts, both philosophical and ontological. [(4a)-(4c)]

³⁷⁴ By ‘ontological representations’ I am including here representations of addiction in important classification systems and vocabularies that are not strictly speaking ontologies, such as the DSM-V and the ICD-11. I explain why below.

Section 5.4 closes with a brief comment about AddictO, a new addiction ontology being developed.³⁷⁵

Section 5.5 concludes the chapter by reviewing the support that has been garnered thus far for the conclusion of the overarching argument of my project.

5.2 Evaluation Criteria

This section lays out the criteria by which the competitor accounts and representations of addiction will be evaluated. I start with the substantive criteria, which I argue ought to be captured by any good account of addiction. Some were introduced in the preceding chapter's discussion about the key addiction-relevant phenomena. In addition, the substantive desiderata include:

- (i) the assumption I started my project with (there are addicts);
- (ii) two of the three desiderata from Chapter 1 (empirical sensitivity and proper level of analysis),³⁷⁶ and,
- (iii) two components of the account defended in Chapter 3 (desires and impaired control).

Next, I turn to the methodological criteria. Any good ontology ought to capture these when representing addiction (and ultimately any entity), and all are drawn from the principles of best practice underlying BFO and discussed in Chapter 2.

While I have already discussed most of the below criteria, including why they are worth capturing, I offer brief explanations of the criteria as a reminder of the general idea behind them. Those we have not seen will be explained and defended as needed. Let us turn now to the criteria.

5.2.1 Substantive Desiderata: How to Conceptualize Addiction

I start by simply providing the list of substantive criteria in Table 2 below. Brief corresponding explanations and, where applicable, defenses of these criteria appear after the table.

Table 2: Substantive Desiderata (Account of Addiction)
1. There are addicts (as distinct from non-addicts)

³⁷⁵ See Hastings et al. (2020).

³⁷⁶ The third, *theoretical breadth*, is captured by the inclusion of some of the other criteria.

2. Substance and behavioral addictions are subtypes ³⁷⁷
3. Addiction is distinct from its causes and effects
4. Addiction is distinct from dependence
5. Addiction is distinct from addicted behavior
6. Exclude a harm condition from the definition ³⁷⁸
7. Exclude an historical condition from the definition
8. Allow for the possibility of willful addicts
9. Exclude disease from the definition
10. Ensure the account is empirically sensitive
11. Ensure the account has the proper level of analysis
12. Addiction involves desires and impaired control
13. Addiction is, at a minimum, a disposition ³⁷⁹

Table 2: *Substantive Desiderata (Account of Addiction)*

First, *there are addicts and there are non-addicts, they are distinct, and addiction is the difference*. This is just a restatement of the assumption my project started with. I refer the reader to Section 1.1.1 in Chapter 1 for the motivation and explanation behind this assumption.

Second, *substance and behavioral addictions are subtypes (of addiction)*. Any account of addiction should capture the fact that **SUBSTANCE ADDICTION** (for which Jim’s heroin addiction would be an instance) and **BEHAVIORAL ADDICTION** (for which Kenny’s gambling addiction would be an instance) are different subtypes of **ADDICTION**. In other words, whatever it is that makes something an addiction is also what makes a substance addiction an addiction (to some type of substance) and a behavioral addiction an addiction (to some type of behavior).³⁸⁰ By analogy, consider providing an account of what a sport is that fails to capture the fact that **SOCCER** is *subtype_of* **SPORT**.

³⁷⁷ This does not imply that any instance of behaviors that are commonly referred to as ‘behavioral addictions’, such as gambling addiction, sex addiction, and shopping addiction, are always genuine addictions (as opposed to just addiction-like). The point here is much simpler (almost trivial). For any behavior that is a *genuine* addiction, that behavior will be an instance of the more general type, **ADDICTION**. It will become clear below why this is included.

³⁷⁸ Note that this does not say that addictions cannot be harmful, nor that addiction cannot be understood as likely (even very likely) to produce harm. The same is true for other criteria that are meant to *exclude* particular features from appearing in the definition of ‘addiction’, such as an historical condition or disease.

³⁷⁹ The claims that desires, control, a dispositional nature, and so on are required of a good account is not assuming my thesis from the start. The preceding chapters defended the inclusion of these particular features at length.

³⁸⁰ Moreover, this criterion would still apply even if one thought there was no genuine difference between substance and behavioral addictions (a view I am inclined to accept).

Third, *addiction is distinct from its causes and effects*. Any account of addiction should capture the fact that addiction is not the same thing as what causes or is caused by it. This should seem obvious, and perhaps even trivial. However, while I agree, the state of the literature is such that this claim cannot be seen as trivial. Violation of this criterion occurs in different forms in the literature, demonstrating how researchers can conflate one or more of the causes or effects of an addiction with the addiction itself. Four versions of this mistake seem to occur more often, and hence are worth focusing on specifically. Thus, the next four criteria are essentially more specific versions of this one.

Fourth, *addiction is distinct from dependence*. Any account of addiction should capture the fact that addiction is not identical to dependence. Dependence is tolerance and the disposition to withdrawal.³⁸¹ It is possible to experience dependence without addiction and addiction without dependence.³⁸² While dependence may often lead to addiction, this makes them neither identical nor conceptually connected in a way that would justify defining one in terms of the other.

Fifth, *addiction is distinct from addicted behavior*. Any account of addiction should capture the fact that addiction is not identical to the particular addicted behaviors that someone with an addiction engages in *qua* addict. Not all behaviors of an addict are *addicted behaviors* – that is, realizations of the addiction that is present – and so there is some important distinction between an addict’s *behavior* and their *addicted behavior*. Nonetheless, the addiction is not the behavior that it manifests. This is most easily seen in the fact that addicts remain addicted when they are asleep and not behaving at all.

Sixth, *exclude a harm condition from the definition*. Any account of addiction should capture the fact that addiction is not necessarily harmful, even if harm is a common consequence of being addicted. A possible effect of failing to control one’s desire to engage in certain types of behaviors, especially when

³⁸¹ I am using ‘dependence’ here in the standard way it is understood in the literature.

³⁸² Consider, for instance, pain patients that become dependent yet stop using as soon as their treatment is over (Ballantyne & LaForge, 2007, p. 249). That dependence and addiction come apart is also widely accepted (George Koob, 2001; Ralphs et al., 1994; Savage et al., 2003). The distinction helps to make sense of such cases where tolerance and withdrawal come apart from an impaired ability to control the relevant behavior, and it is why the DSM-V dropped the DSM-IV’s use of ‘dependence’ for diagnosing addiction (O’Brien, 2011).

related to alcohol, drugs, gambling, and the like, is that some amount of harm will occur. However, this does not entail that addiction is necessarily harmful. There is nothing incoherent about a person being addicted yet suffering no harm *in virtue of their addiction* (perhaps due to vigilance, luck, isolation). Moreover, risk of harm is not harm. Hence, ‘addiction’ should not be *defined* in terms of harm.

Seventh, *exclude an historical condition from the definition*. Any account of addiction should capture the fact that addiction is distinct from any particular causal history that brings it about. Perhaps repeated use of a particular substance is a *primary* cause of an addiction to that substance. However, this would not entail that ‘addiction’ should be *defined* in terms of being caused by repeated use of that substance or any other. It is simply one (perhaps common) way in which an addiction could have been caused to exist. *Requiring* an historical condition makes an account susceptible to counterexamples involving, for instance, instant addicts (discussed in the previous chapter), or a causal history of a different type than required by the account. What causes *X* is distinct from *X* (setting aside self-causing *X*-s). Hence, addiction should not be defined in terms of how it was caused to exist.

Eighth, *allow for the possibility of willful addicts*. Any account of addiction should capture the fact that addiction is not necessarily ambivalent. That is, it is possible for an addict to wholeheartedly endorse his addiction. Again, the main point here is to keep what is common or typical distinct from what is essential or necessary.

Ninth, *exclude disease from the definition*.³⁸³ Any account of addiction should capture the fact that addiction is possibly not a disease. That is, it seems unclear whether every possible instance of addiction must be an instance of a disease. First, both the disease status of addiction and definitions

³⁸³ This *does not* imply that instances of addiction cannot be diseases (or, if you like, dysfunctions). Keeping disease (or dysfunction) out of the definition simply means that we are not requiring that any possible instance of addiction is necessarily an instance of disease (or dysfunction). This in no way entails that *no* instance of addiction can be a disease or dysfunction. This would be like saying that someone who excludes *having a back* from the definition of ‘seat’ (so they can count bar stools as seats, for instance) would be ruling out seats ever having backs. This is an invalid inference. Seats could still have backs even with this exclusion. The exclusion simply says that seats do not *necessarily* have backs since *having a back* is excluded from *what it is* to be a seat. This claim does not even entail that some actual instance of addiction is not an instance of disease (or dysfunction). Again, the exclusion criteria facilitate neutrality on such points.

of ‘disease’ are far from settled questions in the literature. For this reason, and given the value of achieving unification, we ought to refrain from defining ‘addiction’ in terms of disease. Second, our understanding of the nature of addiction may be biased towards unhealthiness and dysfunction due to the fact that much of our data on addiction and addicts comes from clinical and correctional contexts.³⁸⁴ Third, depending on one’s account of disease, it will likely be possible to have two doppelgangers whose different histories result in their having conflicting disease statuses.³⁸⁵ In this case, given the presence of the relevant qualities and dispositions, we should maintain that each is still addicted, but that one addiction is dysfunctional while the other is not. This would entail that addiction is not a disease by definition.

Eleventh, *ensure one’s account is empirically sensitive*. This is just the first desiderata introduced in Section 1.3.3 in Chapter 1. Roughly, it requires that an account of addiction:

- (i) is *consistent* with well-established data;
- (ii) *explains* or is consistent with the best explanation(s) of the relevant phenomena, such as the puzzle of addiction; and,
- (iii) is *neutral* with respect to any debate or point of controversy where the evidence is less than conclusive, and controversy remains.

Twelfth, *ensure one’s account has the proper level of analysis*. This was the third desiderata introduced in Chapter 1. Roughly, the idea is that an account of addiction should adopt a domain-neutral approach as far as possible and avoid focusing too specifically on a particular disciplinary perspective. It is also about the ability to unify different accounts that utilize distinct approaches or perspectives.

Thirteenth, *addiction is, at a minimum, a disposition*. Any account of addiction should capture the fact that addiction and its surrounding phenomena are best accounted for when the former is

³⁸⁴ See Heyman (2013).

³⁸⁵ For instance, if evolutionary history counts towards some condition being a dysfunction (and so a disease), then it will likely be possible for two qualitatively identical individuals with slightly different evolutionary histories to differ with respect to whether some particular dysfunction is present. In such a case, one would be an addict and one would not, despite having exactly the same kind (and strength, etc.) of disposition towards desires and impaired control, exactly the same kind of brain makeup, and so on. I am not claiming that the differing *disease* statuses is a problem. I am claiming that both should be considered addicts, but if addiction is necessarily a disease, they would not be.

understood as a disposition. The preceding chapters partly served to defend this claim. The dispositionalist account best explained key facts about addiction and is consistent with what we know about it empirically. It also unifies almost all other existing accounts and helps to avoid a number of counterexamples. Consider whether you think an addict is still an addict when they: go to sleep; are not using or even thinking about using; are suddenly dropped on a desert island with no object of their addiction; suddenly slip into a coma; are in the first few days of their recovery; and so on. If “yes” is a plausible answer to these and similar questions, then the most plausible account of addiction ought to begin with its being some type of disposition. I provided a particular account of the nature of this disposition in preceding chapters. However, the criterion here is less demanding. It only requires that a good account of addiction start with its being a disposition *of some kind* and can then fill in or remain agnostic about the further details.

Let us now turn to the methodological criteria concerning ontologizing addiction.

5.2.2 Methodological Desiderata: How to Represent Addiction in an Ontology

I again start by simply providing the list of methodological criteria in Table 3 below. Brief corresponding explanations appear after the table. There are further principles of best practice beyond those in the Table 3 that an ontology must follow (Appendices B and C). However, I focus on these because they were most often violated by the existing ontologies representing addiction.

Table 3: Methodological Desiderata (Ontology)	
1. Definition is present	1a. Non-circular
	1b. Complex to simple
	1c. Aristotelian form
2. Consistent backbone is_a hierarchy is present	2a. Definition of “is_a”
	2b. Transitivity
	2c. Single inheritance
	2d. Sparse ontology
3. Follow naming conventions	3a. Singular nouns
	3b. Provide clarity

	3c. No logical operators
4. Conform to realist methodology	
5. Ensure ability to link up to BFO	

Table 3: Methodological Desiderata (Ontology)

First, *a definition must be present*. Any good ontology must provide a definition for each class that it includes. Without a definition, that class (and the ontology in general) loses much of its utility. It is possible to maintain a type-subtype hierarchy without a definition of some term, but this would still require that higher-level parent classes are properly defined. Otherwise, even the hierarchical structure becomes meaningless and lacks any real utility.

The three sub-components of the first criterion are as follows. First, *the definitions should, as far as possible, be non-circular*. That is, the term being defined should not appear in the definition.³⁸⁶ Ideally, terms should also avoid circularity by not being defined in terms of each other where no further information is given. A toy example is defining ‘chair’ as *a seat* and ‘seat’ as *a chair*. Second, *definitions should include either simpler terms than the term being defined or terms that are already defined elsewhere in the ontology*. The latter may happen for definitions of technical terms. The main idea is that definitions should make the term being defined easier to understand, not harder. Third, *definitions must be in Aristotelian form – A is_a B that C*. Roughly, this ensures that a consistent backbone hierarchy of *is_a* relations can be created and provides a basis for expanding that hierarchy and distinguishing the classes therein from their parents, children, and siblings.³⁸⁷ This leads to the next criterion.

Second, *a consistent backbone is_a hierarchy must be present*. Any good ontology must have a hierarchically structured taxonomy of type-subtype (*is_a*) relations serving as its backbone.³⁸⁸ This should follow straightforwardly from successful adherence to the principle requiring Aristotelian

³⁸⁶ I am using ‘definition’ here in a more colloquial sense, as in *unmarried eligible male* is the definition of ‘bachelor’. This is in contrast to ‘definition’ referring to the combination of a *definiendum* (that which is being defined) and the *definiens* (that which is used to define the *definiendum*). Clearly the term being defined must occur in the definition when used in the second way. If you prefer the more technical sense, the rule becomes: keep the *definiendum* out of the *definiens*.

³⁸⁷ Chapter 2 contains a fuller explanation of the importance of Aristotelian definitions for ontology.

³⁸⁸ Again, the reader is referred to Chapter 2 for a fuller discussion of why this is important.

definitions. More than this, though, the ontology should utilize a *consistent* hierarchy of *is_a* relations, such that the four subcomponents of this criterion are met. Additionally, while it is not specified as its own subcomponent, a consistent *is_a* hierarchy also entails that no class is defined or classified in such a way that inconsistent types get mixed. A simple example that will be important below is that anything which *is_a* **CONTINUANT** should never be defined or classified as a *subtype_of* **OCCURRENT**, nor vice versa. These are different kinds of entities with incompatible essential properties and relations. This would be like classifying **PROCESS** as a *subtype_of* **MATERIAL OBJECT**, or vice versa.

The four subcomponents of the second criterion are as follows. First, *the hierarchy should respect the definition of the is_a relation*. ‘**A is_a B**’ means that every instance of **A** is also an instance of **B**. Hence, if **HUMAN is_a ANIMAL**, then everything that is an instance of the type **HUMAN** is also an instance of the type **ANIMAL**. Moreover, the same applies to the *differentia* (the ‘that Cs’ portion of the Aristotelian definition), since these should be essential features specifying what it is that makes something an instance of that type and not some other subtype of the parent class. Thus, if we further specify that **HUMAN is_a ANIMAL that is rational**, then every instance of **HUMAN** will also be rational (in addition to being an instance of **ANIMAL**), as will any child classes of **HUMAN**.

Second, *the is_a relation is transitive such that if A is_a B and B is_a C, then A is_a C*. Hence, if **HUMAN is_a ANIMAL** and **ANIMAL is_a ORGANISM**, then **HUMAN is_a ORGANISM**. Moreover, this transitivity also means that children inherit *all of the essential properties* of their parent classes. So, if metabolizing is essential to being an *instance_of* **ORGANISM**, then every human and animal will also metabolize. The main idea is that, once an *is_a* hierarchy is constructed, essential features of a class – their parent type and their associated *differentia* – are inherited downward.

Third, *no class should have more than one (direct) parent*. In our example, humans are both animals and organisms, but only **ANIMAL** is the direct parent of **HUMAN**. Thus, if **A is_a B**, then **A is_a C**

only if **C** is a parent of **B** (or of **B**'s parent, or **B**'s parent's parent, and so on) and there is no other violation of single inheritance.

Fourth, *a class should not be asserted that combines entities through conjunction or disjunction*. For instance, there should not be classes like **WATER OR DRINKING**, **DOCTORS AND SURGERIES**, and so on. The transitivity of the *is_a* relation makes such ontological assumptions problematic because it allows types (and instances) with inconsistent properties and relations to be lumped together, disrupting the reasoning power we want to get out of our ontologies.³⁸⁹

The third methodological criterion is: *naming conventions are followed*. Failing to follow the proper conventions for naming or labeling classes undermines the intuitiveness and understandability of the ontology. Both help to ensure the ontology is maximally useful and interoperable.

The subcomponents of this third criterion are as follows. First, *terms should be singular nouns or noun phrases, as opposed to pluralized ones*. For instance, use 'human' instead of 'humans', 'intentional act' instead of 'intentional acts', and so on. Second, *ontologies require understandability as far as possible, and so terms should also be clear*. In line with definitions moving from the complex to the simple, labels should be intuitive, easy to grasp, and in general provide more clarity than confusion. Third, *labels should, whenever possible, avoid logical operators such as 'and', 'or', and the like*. This is in line with the above-mentioned prohibition on introducing classes through conjunction or disjunction.

The fourth criterion is: *conform to the realist methodology*. Any good ontology must be realist, such that its class terms are intended to represent entities in the world and not our ideas, knowledge, or concepts of such entities. For instance, 'human' should not be defined as *the concept of a person as a*

³⁸⁹ Just consider how it sounds to classify types in this way. Suppose we included the class **DOCTORS AND SURGERIES** in our ontology, and then classified both **OPEN HEART SURGERY** and **PEDIATRICIAN** as *subtype_of* this class. First, it sounds very odd to say, "A pediatrician is a kind of doctor and surgery." Second, where would we classify **DOCTORS AND SURGERIES**? Under medical role? Under medical process? Does "Doctors and surgeries are a kind of medical process" sound any better? Third, we would have to combine the essential properties of both **DOCTOR** and **SURGERY** as the *differentia* for this class, which would mean that any subtype (such as **PEDIATRICIAN**) would inherit all of these properties (such as the essential properties for being a surgery). This is all problematic both metaphysically and practically.

rational animal or some such thing. Similarly, classes such as **OBSERVED CELL**, **UNIDENTIFIED AIRCRAFT**, or **DISEASE THAT CAN BE MEASURED** should be avoided, since these refer to *our access* to these entities as opposed to the entities themselves.³⁹⁰

Fifth, and finally, *ensure the ability to link up to BFO*. Any good ontology must be such that, for any class in the ontology, there should be a clean (single, consistent) chain of *is_a* relations between that class and a class in BFO. This might be by having a BFO class as the root node in the ontology, as in the Process Ontology (PO). Every class in PO is a child (directly or by inheritance) of BFO **PROCESS**. Alternatively, the root nodes in the ontology may simply have a clean connection to a BFO class. At a minimum, though, every class in the ontology needs to *ultimately* link up (through some series of *is_a* relations) to the appropriate BFO class. The basic assumption here is that BFO and its underlying principles are the proper standard for building ontologies.

With these two sets of criteria on the table, we can now turn to the evaluations of the extant views and ontological representations of addiction in the literature.

5.3 Evaluating Extant Accounts of Addiction

In this section, I evaluate a number of prominent accounts of addiction. Many should be familiar from preceding chapters. There is a seemingly unending list of researchers working on addiction, each with either their own view or their own take on the specifics of a shared, more general view. Thus, it would be impossible to survey every version of every type of account on offer (let alone every particular account on offer). The selection of views here was intended to provide some diversity in both the type of account as well as the background of the author defending it. There are disease views (both brain-based and non-brain-based), non-disease views, *no difference* views arguing that

³⁹⁰ There are workarounds for capturing such things. For instance, there might be the classes **AIRCRAFT** and **IDENTIFICATION PROCESS**, as well as a relation that links these two together. In this way, we could assert that some **AIRCRAFT** was *participant_in* some **IDENTIFICATION PROCESS**, while another was not, without a non-realist class.

addiction and ordinary *akratic* action are not different in kind, psychology-based views, neuroscience-based views, decision-theory-based views, philosophical views, and so on. In this way, while I cannot address every view (nor every *type* of view), the pattern that will emerge will be representative of the literature more generally.

On a similar note, I want to address a potential worry that may arise from consideration of the substantive criteria that I am using as the basis of my evaluation. Addressing this worry entails a spoiler (though, one that was probably already anticipated). The dispositionalist account – in particular, the one I have been defending – is going to score well on these criteria, while the other accounts surveyed will fail on one or more of them (usually more). One might worry that it is a bit convenient that the criteria I provide just happen to be well-suited to a dispositionalist account (especially the thirteenth, requiring that a good account be, at a minimum, dispositionalist). Let me address this worry.

Most of the criteria – essentially all of them except the thirteenth – are supported independently of the dispositionalist account. For instance, that an account must capture the fact that **SUBSTANCE ADDICTION** and **BEHAVIORAL ADDICTION** are each a *subtype_of* the more general class **ADDICTION** is completely unrelated to whether that account is dispositionalist or not. The same goes for addiction being distinct from what causes it and from what it causes. These are simple, intuitive claims about metaphysics more generally. Setting aside self-caused entities (which addiction is not), if X is caused by Y , then Y is not the same thing as X .³⁹¹ This alone accounts for roughly half of the substantive criteria. What is left is defended in a similar way. In other words, the first twelve substantive criteria are not inferred on the basis of the last. The opposite is true: the thirteenth criterion

³⁹¹ This is true even if the causation is synchronic. Your sitting down on the seat cushion might occur at the same time as the indentation (shaped like your rear end) forming in the seat cushion. Still, your sitting down and the indentation forming are not necessarily the same thing, but are rather co-occurring events. If they were the same thing, the event of the indentation forming would have caused itself. What is more, I doubt very much that we have to worry about such things since the causes and effects of addiction are probably not going to be co-occurring with addiction. Moreover, the causes and effects of addiction that are relevant to the evaluations (because other views include them) are not co-occurring, such as *repeated use* and *harmful consequences*. First someone repeatedly uses, and then the addiction forms. First an addiction forms, and then harmful consequences (might) ensue.

is inferred from thinking about the first twelve, along with the defense of the dispositionalist account throughout the preceding chapters. Thus, it is no accident that the criteria a good account should meet are well-suited to a dispositionalist account. This is precisely what we should expect after considering what we know about addiction, what is intuitive about addiction, what best unifies the phenomena and existing accounts, and so on. None of this is question-begging or otherwise illicit. It is simply where we ended up after careful consideration of the relevant evidence.

Now to the evaluations. I consider each view in alphabetical order according to the author(s) associated with the account. Many views and authors should be familiar from earlier chapters. A full exploration of the intricate details of each view is beyond the scope of my project. Thus, I will simply provide evidence for the main tenets of their view and identify which criteria are violated.

5.3.1 George Ainslie

Ainslie's account focuses on the choice patterns of addicts, famously incorporating the notion of *temporal* or *hyperbolic discounting* in his account.³⁹² The details of hyperbolic discounting are not important for our present purposes, which is just to evaluate Ainslie's account *vis-à-vis* the substantive criteria above. What is relevant about this sort of account, though, is that it focuses heavily on individuals' actual choices and behaviors. With this in mind, let us turn to evaluating Ainslie's account, the core of which can be summarized as follows:

addiction: *the habitual overvaluation of the present moment that an agent fails to counteract in a specific area of their life over some period of time.*³⁹³

First, Ainslie's account violates substantive criteria 3 and 5 because he does not keep addiction distinct from its effects, namely, the behavior and choice patterns it can produce.³⁹⁴ While 'habitual'

³⁹² See Ainslie (2001, 2017, 2019). Again, Appendix A discusses hyperbolic discounting and its relation to addiction.

³⁹³ See Ainslie (2019, p. 37). This is not a quote. Instead, it is my own summary of Ainslie's core understanding of what addiction is based on what he says in the foregoing reference and elsewhere. In this and the following subsections on extant accounts, I will use italics indented (and hanging) in this way when providing the authors' definitions of 'addiction'.

³⁹⁴ Henceforth, I will use 'SC' and number of the criteria violated, such as 'SC3' and 'SC5'.

suggests a dispositionalist account, Ainslie includes the idea that addiction is about actually failing to counteract their overvaluation of the present moment. In other words, he focuses on *actual patterns* of choice and behavior. On his view, addiction seems to be the *exercise* of a disposition towards exaggerated delayed discounting as opposed to the disposition itself. Consider what he says in the following two quotes (emphases added):

So the problem for the science of addiction is not an addict's susceptibility to temptation, but *why she fails to use her culture's shared knowledge to counteract it* in specific areas over part of her life.³⁹⁵

Addictions are not simply recurrent impulses, but *complex compromises with your long-term interests* that develop when you *try repeatedly to resist a temptation and fail*.³⁹⁶

Second, as the second of these quotes shows, Ainslie's account also violates SC7 that prohibits an historical condition. This is because it includes that addiction must develop from repeated failures to resist temptations.

Third, Ainslie's account implies that addiction is necessarily ambivalent, violating SC8 that requires the possibility of willful addicts. In fact, Gene Heyman has an Ainslian account and explicitly requires ambivalence. This is because such accounts are grounded in the idea that addiction is, roughly, exaggerated hyperbolic discounting, and thus that there is always some long-term, competing motivation to *not* use to which the addict gives in at the moment of choice.

Fourth, that Ainslie's account focuses squarely on the addict's choice patterns – that is, their temporal discounting in moments of conflicting desires – pushes it towards a violation of SC11. This criterion requires that an account have the proper level of analysis such that it can incorporate (and, ideally, unify) different disciplinary perspectives and approaches.

Fifth, Ainslie's focus on actual patterns of choice and behavior also violates SC13 as well, requiring that addiction be, at minimum, a disposition.³⁹⁷

³⁹⁵ Ainslie (2019, p. 37).

³⁹⁶ Ainslie (2017, p. 237).

³⁹⁷ One might wonder two things at this point. First, I am saying that addiction is the disposition and Ainslie is focusing on the behavioral manifestation (and others are focusing on other aspects of the disposition like the material basis and

5.3.2 Bennet Foddy and Julian Savulescu

Foddy and Savulescu defend what they call the “liberal account of addiction,” defining ‘addiction’ (and ‘appetite’, the main component of the definition) as follows:

addiction: *a strong appetite, which is a disposition that generates desires that are urgent, oriented toward some rewarding behavior, periodically recurring, often in predictable circumstances, sated temporarily by their fulfillment, and generally provide pleasure.*³⁹⁸

Their account is liberal because, given what they mean by ‘appetite’ (and their failure to say what they mean by ‘strong’), addiction ends up being no different than *akrasia*, which almost everyone experiences at least sometimes. Hence, the account is liberal in the sense of being much more permissive than most others – probably everyone is addicted to something (or many things) on their view, such as coffee, water, sex, tennis, pie, television, love, seeing family, and so on. Part of Foddy and Savulescu’s motivation is that there is insufficient evidence for the claims that addicts value satisfying their addictive desires above anything else, and that addicts cannot behave autonomously. Hence, they end up concluding that *addictive desires are just strong, regular appetitive desires.*³⁹⁹

First, on one plausible interpretation of the liberal account, Foddy and Savulescu violate SC1, the fundamental assumption that there are addicts and non-addicts, and something makes this difference. Note that this assumption *does not* require that the difference be a genuine difference in kind. If the difference ends up being a matter of degree (or along the lines of a defined class), this is enough to satisfy SC1. Still, because of the ease with which this definition is satisfied and the lack of an explanation for what ‘strong’ comes to, it is unclear how these authors can distinguish between

external triggers). So, why think the disposition is the addiction and not, for instance, the manifestation as Ainslie suggests? The answer consists in the arguments in the preceding chapters, such as that the dispositional account best explains the phenomena and also unifies existing accounts. Second, could Ainslie not just reconceptualize his account as the *disposition* towards certain choice and behavioral patterns (including hyperbolic discounting)? The answer here is, yes, he should. That is what I have been arguing. Addiction is a disposition, not the manifestations of that disposition. Were Ainslie to reconceptualize his account to make it a disposition, he would be accepting one of my main conclusions of the dissertation and, in my view, on a much better track to correctly understanding addiction.

³⁹⁸ See Foddy & Savulescu (2010b, p. 35). Their concise definition is *a strong appetite*.

³⁹⁹ Foddy & Savulescu (2006, p. 14).

addicts and non-addicts. Hence, a reasonable way to read their account is that, at bottom, they think there really is no difference between addicted and *akratic* behavior. Since the latter is standardly thought to be non-diseased, non-compulsive, free, and responsible behavior, it is reasonable to conclude that Foddy and Savulescu are nihilists about addiction. For various reasons discussed and defended throughout the preceding chapters, I think this is an untenable position. Still, they might capture SC1 by appeal to the difference in *strength* referred to in the definition. However, this difference has yet to be worked out in any detail by the authors.⁴⁰⁰

Second, Foddy and Savulescu's definition implies that addicted behavior must be present, confusing addiction for some of its particular effects and thus violating SC3 and SC5. This is because they say that an appetite generates desires that are "periodically recurring." Hence, despite their laudable appeal to dispositions, the authors' definition mistakenly requires that the disposition must recurrently manifest. This belies the fact that the authors are likely conflating the disposition with its manifestations and not fully taking the dispositionalist framework on board since, were they to really count addiction as a disposition, there would be no need to require periodic recurrence.

Third, this account violates SC10, requiring that the account be empirically sensitive. In particular, it violates the *neutrality* component of this criterion since, as noted above, Foddy and Savulescu seem to contend that there is simply no difference between addiction and ordinary *akrasia*. Thus, while their account may be sensitive to the data in other ways, it commits itself to a highly controversial position.⁴⁰¹

Fourth, this account also violates the unifying component of SC11. Since the liberal account virtually does away with any difference between addiction and ordinary *akrasia*, their account will have

⁴⁰⁰ Moreover, were it to be worked out, it seems like a good way to do this would be to appeal to differences in the strength of the dispositions of addicts relative to *akratics*. This is precisely how the dispositionalist account I defend would handle this, and it would start to (correctly) bring Foddy and Savulescu's account into the dispositionalist framework.

⁴⁰¹ Even most cotemporary choice theorists accept that there is a difference between addicts and non-addicts (Heyman, 2009; M. Lewis, 2015; Sinnott-Armstrong & Pickard, 2013). This reading of Foddy and Savulescu puts them somewhat on a par with the outlier (and outdated) moral model of addiction with respect to the issue of control.

an extremely hard time being consistent with, let alone incorporating or unifying, the plethora of extant views that do recognize such a difference.

Fifth, and finally, Foddy and Savulescu's account violates the impaired control component of SC12, requiring that an account of addiction capture the fact that it impairs control to some degree. One might think that *akrasia* is a form of impaired control, which seems plausible, and so the account *can* capture this criterion. However, the authors explicitly argue against the claim that addiction impairs control, and so this interpretation is not available to them.⁴⁰²

5.3.3 Nick Heather

Like Ainslie, Heather focuses on the behavioral aspects of addiction as they relate to the addict's competing motivations and commitments to changing or ceasing their addicted behavior. However, like Foddy and Savulescu, Heather also sees addiction as simply an extreme form of *akrasia*.⁴⁰³ Moreover, as noted in Chapter 1, Heather has voiced his concerns with the state of the literature, specifically concerning the lack of clarity surrounding definitions and characterizations of addiction. Accordingly, he offers a straightforward definition of 'addiction' as follows:

addiction: *a repeated and continuing failures [sic] to refrain from a specified behavior despite prior resolutions to do so.*⁴⁰⁴

First, as with Foddy and Savulescu, Heather's definition risks violating the fundamental assumption in SC1 that there are addicts *as distinct from non-addicts*.⁴⁰⁵ Again, this is because everyone experiences *akrasia* and so, if addiction is simply *akratic behavior*, Heather is forced to either violate SC1, hold that everyone is addicted, or put forward an account of what 'extreme' comes to that can

⁴⁰² Their argument also rests on poor evidence for a *no impaired control* condition, such as that purported addicts tend to age out of their addictions without treatment and that purported addicts seem to make controlled choices sometimes (like not smoking on a plane). Neither entails that addiction fails to impair control, especially when understood in the dispositionalist sense that the account defended here offers.

⁴⁰³ See Heather (2017a, 2020).

⁴⁰⁴ See Heather (2017a, p. 147).

⁴⁰⁵ This is not about the possibility of universal addictions, which is consistent with my dispositionalist account. The point is about collapsing the distinction between addicts and non-addicts (such as experiencing one-off *akratic* acts).

adequately distinguish *ordinary akrasia* from *akrasia qua addiction*. He has not done so, and only suggests that addiction is a “form with more severe negative consequences than, but nevertheless on a continuum with, what might be termed ordinary akrasia.”⁴⁰⁶

Second, Heather’s definition violates SC3 and SC5, which require that addiction be kept distinct from its effects and behavioral manifestations. Addiction, the reader will recall, can be present in the absence of performed addicted behaviors. In contrast, Heather identifies addiction with *repeated and continuing failures*, indicating that addiction is identified with those behavioral failures and not, more plausibly, the disposition towards such failures.

Third, as the above quote shows, Heather violates SC6 that requires addiction not be defined in terms of its harmful consequences. He claims there that harmful consequences are the part of the definition that help to distinguish addiction from ordinary *akrasia*.

Fourth, understanding addiction as a form of *akrasia* entails that Heather’s definition violates SC8 as well. This is because *akrasia* requires that an individual acts against their better judgment (“despite prior [contrary] resolutions...”). Hence, Heather’s definition does not allow for the possibility of willful addicts. Here is Heather on addiction meeting the requirements of *akratic* action (emphasis added to the relevant clauses):

...addiction conforms to four definitional requirements of akratic action, that it must be: (i) free; (ii) intentional; (iii) *contrary to the agent’s better judgment based on practical reasoning*; [and] (iv) *consciously recognized as contrary to better judgment* at the time of action.⁴⁰⁷

Fifth, like Foddy and Savulescu’s definition, Heather’s definition violates the neutrality component of SC10 and the unifying component of SC11. The reason is the same. First, an account of addiction that identifies it with *akrasia* commits itself (or, as noted, strongly risks committing itself) to the controversial view that there really is no difference between addiction and what we would

⁴⁰⁶ See Heather (2020, p. 1). Again, it seems that a good way to provide an account of the difference – in terms of either strength of frequency – is by appealing to differences in the dispositions of the agents.

⁴⁰⁷ Heather (2020, p. 1).

normally consider ordinary behavior.⁴⁰⁸ Second, and consequently, this view will have a hard time incorporating different disciplinary perspectives and approaches, especially those from clinical and neuroscientific realms that highlight significant impairments to control, at the least.

Sixth, the definition violates the component of SC12 that requires some level of impaired control. This is seen in the first condition of his view of *akrasia* quoted above. Heather, like many others, holds that *akratic* actions are necessarily free, and hence, controlled. This criterion *does not* entail that addicts never act freely. However, Heather's definition suggests that addicts do not suffer *any* loss of control in virtue of their addiction.

Seventh, and finally, Heather's definition focuses on actual behaviors ("repeated and continuing failures"), and so violates SC13, requiring that addiction is minimally a disposition.

5.3.4 Gene Heyman

Heyman, a prominent opponent of the brain disease model of addiction, provides an Ainslie-style account that focuses on patterns of choice and behavior and is grounded in hyperbolic discounting. Hence, Heyman's account violates all of the same criteria as his predecessor (SC3, SC5, SC7, SC8, SC11, and SC13), and for the same reasons. His account can be summarized as follows:

addiction: *ambivalent drug use that eventually involves more costs than benefits, or consumption of a highly preferred substance or activity that is: (i) ambivalent, (ii) excessive, and (iii) persistent despite being on balance more costly than beneficial.*⁴⁰⁹

In addition, though, Heyman violates two further criteria. First, he violates SC6 because he defines 'addiction' in terms of its harmful consequences. Second, he violates the impaired control component of SC12 because he equates addiction to ordinary voluntary action. Here are some

⁴⁰⁸ Of course, as mentioned above, Heather has outs here. However, he has yet to adequately pursue them, and it seems the best out is dispositions in any case. Pointing to more harmful consequences is not going to do the work.

⁴⁰⁹ See Heyman (2009, 2013, 2019).

passages from Heyman that provide textual evidence for all of these violations (emphases added to indicate the relevant clauses):

...addiction is ‘disease-like’ in the sense that it *persists even though on balance its costs outweigh the benefits...* [*it is*] *voluntary behavior* that predicts the persistence of activities that from a global bookkeeping perspective (e.g., long-term) are irrational. That is, addiction is not compulsive drug use, but it also is not rational drug use... the defining features of addiction, which is to say *its destructive and irrational aspects*.⁴¹⁰

...addiction is *ambivalent drug use*, which eventually involves *more costs than benefits*.⁴¹¹

...*voluntary action and addiction differ in degree, not kind*. For example, relapse and attempts to quit using drugs are *signs of ambivalence*, addiction by definition means *excessive drug use*.⁴¹²

5.3.5 Neil Levy

Levy’s account focuses on the neuroscience of addiction and its role in explaining the addict’s preferences at the moment of choice – specifically, the addict’s *shift* in preferences. Recall the Ainslian work on hyperbolic discounting showing that addicts exhibit exaggerated discounting rates as compared to non-addicts, which results in smaller but imminent rewards having much higher expected value than larger but delayed rewards. One way to explain this is that addicts are overcome by their desires at the moment of choice. In other words, addicts genuinely do desire (or value) abstinence, health, and other long-term rewards or goals more. But once the choice to engage in that behavior is imminent, the desire to engage in the addicted behavior ramps up and the addict succumbs.

However, another possible explanation, and one that Levy favors, is that the addict undergoes a shift in beliefs (or judgments) at the moment of choice. That is, at the moment of choice, the addict switches their judgment about what is all-things-considered best. On this view, it is more appropriate to say that an addict is *convinced* of the greater value of using rather than being *overwhelmed* by their desires. Levy explains this at the level of the brain, which he takes to be a *prediction error minimization*

⁴¹⁰ (Heyman, 2013, p. 1).

⁴¹¹ (Heyman, 2013, p. 4).

⁴¹² (Heyman, 2009, p. 124).

machine. According to this view, the function of the brain, and in particular the mesolimbic dopamine system (central to motivation and reward learning), is to minimize and correct for prediction errors between one's *expectation* of the world and its rewards (predicted model) and the *actual* world and its rewards (given model) by updating expectations based on input.⁴¹³ Levy dubs this view the 'belief oscillation hypothesis', and the resulting account defines 'addiction' as follows:

addiction: *the inability of the brain's prediction error minimization system to adapt to prediction error normally, resulting in excessive judgment-shifts due to incentive sensitization.*

First, it is worth point out that it is quite hard to pin down Levy's account of *what addiction is*. This is because, as with many others in the literature, Levy focuses mostly on *explanatory* accounts of addicted behavior. In particular, he appeals to neuroscientific and behavioral evidence to explain the puzzle of addiction, the evidence on exaggerated hyperbolic discounting, and so on. Combing Levy's recent papers did not turn up any straightforward statement of what addiction amounts to, aligning with Nick Heather's previously quoted point about the literature's lack of clarity regarding definitions. In a way, then, his account is not properly an account of the nature of addiction so much as how addiction works and why, neurobiologically, addicts behave the way they do. Nonetheless, since his belief oscillation hypothesis shows up in numerous papers on different addiction-related topics, it seems worth addressing.⁴¹⁴

First, Levy's account seems to violate SC8 because it is unclear how willful addicts would be possible if addiction necessarily involves a shift in judgment at the moment of choice. More specifically, Levy sees addicts as having conflicting beliefs about the relative value of use and abstinence, and his theory is a way of explaining how this conflict is resolved when choice is imminent: their judgments shift.⁴¹⁵ To be sure, he does *not* say addiction just is *akrasia*, but this way of explaining

⁴¹³ See Levy (2019, p. 57).

⁴¹⁴ For examples of such papers, see Levy (2006, 2011b, 2013, 2014, 2015, 2019).

⁴¹⁵ See Levy (2011b).

addiction seems to entail that there is some conflict of judgments about engaging in addicted behavior. Hence, SC8 would be violated.

Second, Levy's account violates SC9, which requires that disease be left out of the definition of addiction. While Levy explicitly argues that addiction is not a *brain disease*, his main point here is that addiction is not necessarily *harmful*. Levy holds a *hybrid* account of disease – roughly, disease is a biological dysfunction plus harm or risk of harm. However, he then goes on to argue that the neurobiology underlying addiction is dysfunctional, but addiction does not necessarily entail harm or risk of harm.⁴¹⁶ Hence, Levy's account makes addiction essentially a dysfunction and violates SC9.⁴¹⁷

Third, while Levy's account is empirically sensitive in important ways, it still violates the neutrality component of SC10. First, his theory is in tension with a widely accepted view of the role of dopamine from Kent Berridge.⁴¹⁸ Levy himself explains that:

For Berridge (2007; Holton and Berridge 2013), then, addiction is a pathology of incentive salience and not reward prediction. It does not involve pathological learning; rather it involves pathological 'wanting'.⁴¹⁹

While Levy suggests that some of his broader points are independent of this debate, the particular theory of what addiction comes to is not. Thus, his account entails a particular, disputed neurobiological model of addiction and dopamine. Second, as noted above, Levy's account offers an explanation of why addicts engage in exaggerated hyperbolic discounting – their judgments shift rather than being overwhelmed by ramped up desires. This stems from his view of *akrasia* as *judgment shift*, which is itself a commitment to a particular theory of what is going on in *akratic* action.⁴²⁰

⁴¹⁶ This argument (Levy, 2013) comes as a reply to an earlier claim to the contrary by a prominent brain disease model proponent (Alan I. Leshner, 1997).

⁴¹⁷ For my purposes, the dysfunctional component of this sort of Wakefieldian hybrid account of disease is all I mean by 'disease' when I say that 'addiction' should not be defined in terms of disease. This was hopefully clear from the reasons provided for this criterion. This is also why I have a separate criterion related to keeping *harm* out of the definition.

⁴¹⁸ See Robinson & Berridge (1993).

⁴¹⁹ See Levy (2014, p. 346).

⁴²⁰ Consider that this is likely the result of Levy attempting to explain *addicted behavior* rather than *what addiction is* (Levy, 2011b). Hence, either his theory violates this criterion in the stated ways, or he has not given us a theory of addiction at all. Neither is a preferable result.

Fourth, Levy's account is narrowly focused on dispositions of *the brain*, and in particular of the role of dopamine, and so violates the unifying component of SC11. Moreover, despite focusing on the brain, Levy straightforwardly pits himself against the widely held brain disease model. Adding to this his idiosyncratic view about why addicts hyperbolically discount delayed rewards, it is unclear how he would be able to incorporate or unify other accounts.

5.3.6 Marc Lewis

In his recent book, *The Biology of Desire: Why Addiction Is Not a Disease*, Lewis argues against the brain disease model of addiction and in favor of a non-disease, learning-based model of addiction.⁴²¹ He does not deny that addicts undergo systematic changes to their brains. Instead, he argues that these changes are insufficient for disease and are only indicative of the addict having formed (that is, *learned*) a habit through ordinary mechanisms of motivation and reward operating as they were designed to.

Here is Lewis on addiction:

So, what exactly is addiction? It's a habit that grows and self-perpetuates relatively quickly, when we repeatedly pursue the same highly attractive goal. Or, in a phrase, motivated repetition that gives rise to deep learning.⁴²²

Again, 'addiction' doesn't fit a unique physiological stamp, like 'disease'...It simply describes the repeated pursuit of highly attractive goals when other goals lose their appeal, plus the brain changes that condense this cycle of thought and behaviour into a well-learned habit.⁴²³

From this, we can summarize Lewis's view of addiction as follows:

addiction: *Habituation to a rewarding behavior that is acquired through motivated (that is, desire-based) repetition that becomes entrenched.*

By 'entrenched' (or 'deep learning' above), Lewis has the addict's neurobiology in mind. Specifically, he is referring to the stabilization of brain pathways that lead to particular types of behavior, which

⁴²¹ See Lewis (2015).

⁴²² Lewis (2015, p. 173).

⁴²³ Lewis (2015, p. 164).

also become more cue selective as habits are established.⁴²⁴ There is a lot to like about Lewis's account and approach, though problems remain.

First, Lewis's account risks violating the fundamental SC1, which requires that we recognize that addicts exist and are distinct from non-addicts. This is not as explicit or clear as it is in, for instance, Foddy and Savulescu's account. However, consider the following from Lewis:

So even though addictive habits can be more deeply entrenched than many other habits, there is no clear dividing line between addiction and the repeated pursuit of other attractive goals, either in experience or in brain function.⁴²⁵

So addiction is not fundamentally different from other unfortunate directions in personality development: a self-reinforcing habit based on intense emotions, encountered repeatedly.⁴²⁶

It is consistent with his account that what Lewis means here is that, despite the absence of a sharp dividing line (or perhaps just our inability to identify it), there are still addicts and non-addicts on either end of the spectrum. The difference between them would be a matter of degree, but this would be enough on my view to meet the requirement of SC1 that addicts exist and can at least be distinguished from non-addicts on the basis of a significant difference on some relevant metric. For instance, Lewis does discuss the extreme difficulty addicts face in changing their behavior and, as noted, the fact that there are systematic changes to their brains. However, given the lack of clarity here, it is worth pointing out that Lewis needs to do more work if he is to meet SC1.

Second, Lewis repeatedly violates SC3, SC5, and SC6 because he conflates addiction with both the events that lead to its development (its causes) and the behavior that it manifests in (its effects). He says addiction develops "when we repeatedly pursue" the same goal. He also says addiction is "*motivated repetition that gives rise to deep learning,*" suggesting that addiction is an entire process unfolding over time that requires particular process parts. Lewis goes on to say that the term

⁴²⁴ See Lewis (2015, pp. 31–32).

⁴²⁵ Lewis (2015, p. 163).

⁴²⁶ Lewis (2015, p. 173).

‘addiction’ “describes the repeated pursuit” of certain goals, as well as changes in the brain. This sentiment is echoed again in the first of the two quotes above. The latter also reinforces the historical condition when he says addiction is “...based on intense emotions, encountered repeatedly.”

Third, the account violates SC11 because, like the brain disease model he targets, Lewis focuses heavily on the brains of addicts. Addiction is not simply a habit in the sense of being a disposition of the person towards certain behaviors. Instead, he focuses on how particular brain regions are changed and says that ‘addiction’ refers to these brain changes in addition to the behavior.

Fourth, Lewis fails to clearly include an impaired control condition, and hence violates SC12. However, it should be noted that this violation is similar to that of SC1 in the sense that the account *apparently* violates this criterion (or at least does not clearly meet it). Thus, perhaps there is some room for Lewis to account for impaired control and thereby a distinguishing feature of addicts that would meet SC1, such as in his discussion of the extreme difficulty addicts face in stopping their behavior. Ultimately, though, this is not presently made clear in his account.

Fifth, Lewis falls short of meeting SC13, the dispositionalist criterion. His talk of ‘habit’ and ‘habitual behavior’ suggest a dispositionalist account is not far from his reach. However, he is not careful to distinguish the disposition itself from the history that produced it, nor from the behavior which manifests it. Hence, we cannot justifiably credit Lewis with a genuinely dispositionalist account.

5.3.7 Hanna Pickard

Pickard has focused on numerous aspects of addiction, such as the social and environmental factors that influence addiction, internal influences such as denial and ignorance, and the *blameless* responsibility of addicts for their addicted behavior.⁴²⁷ Regarding her account on what addiction is, though, Pickard’s view is relatively simple:

⁴²⁷ These are discussed in a number of Pickard’s works (Pickard, 2016, 2017, 2019; Pickard & Ahmed, 2016; Sinnott-Armstrong & Pickard, 2013).

addiction: *a strong and habitual want that significantly reduces control and leads to significant harm.*⁴²⁸

We will not get into the details but suffice to say there is a lot to like about her account, including in particular her emphasis on impaired control, the lack of disease in the definition, and the appeal to a dispositionalist component. Indeed, I will point to only two problems with the account.

First, concerning our criteria, Pickard violates SC6 by requiring that addiction is harmful. She is seemingly aware of the potential for problems here since she addresses the objection that the harm condition is unnecessary or otherwise problematic.⁴²⁹ Her solution is to keep the harm condition because, as she puts it, “diminution of control can arguably count as harmful in itself.”⁴³⁰ The problem here is that Pickard’s explanation of how impaired control is harmful ends up being counterfactual. That is, an apparently harmless addiction is actually harmful since, in some nearby possible worlds, the addict’s impaired control could easily cause them to suffer harm in some way. Thus, on Pickard’s view, impaired control is inherently a risk of harm in some close possible world. This is problematic because, quite simply, a risk of harm is not itself harm. It is at best a chance or likelihood of harm occurring at some later time. Thus, her solution falls flat, and the harm condition remains superfluous.

The second problem is not a direct violation of one of the substantive criteria. As noted, her account has plenty going for it. That said and considering the comparative goal of this chapter, it is worth pointing out that Pickard’s account is rather underdeveloped with respect to the metaphysical nature of addiction. Pickard’s work has focused a lot on what influences an addiction’s development and manifestation, as well as how we should treat addicts (her *responsibility without blame* project⁴³¹). This is important work. Still, it is at least as important to understand the nature of the entity one is interested in if such practical goals are going to be addressed as accurately and comprehensively as possible. In

⁴²⁸ Sinnott-Armstrong & Pickard (2013, p. 861).

⁴²⁹ See, in particular, Sinnott-Armstrong & Pickard (2013, pp. 860–861).

⁴³⁰ Sinnott-Armstrong & Pickard (2013, p. 861). The explanation that follows comes from here as well.

⁴³¹ See Pickard (2017) and also <https://www.responsibilitywithoutblame.org>.

other words, understanding what influences addiction and how we should treat those who have it presupposes an understanding of what addiction is. Understanding these things certainly comes in degrees, but where we fall short on the latter, we will equally fall short on the former. Hence, the account defended here can be seen as supplementing an account like Pickard's by filling in a significant part of the metaphysical story and avoiding pitfalls like the harm requirement.

5.3.8 Nora Volkow

Volkow is a neuroscientist and leading proponent of the brain disease model of addiction. Hence, like Levy and Lewis, she focuses on the neurobiological underpinnings of addiction. Volkow has put forward numerous characterizations of addiction, some of which have slight variations, but her view can be summarized as follows:

addiction: *a dysfunction in the brain that manifests as oversensitivity of the reward/motivation system to stimuli/cues associated with some certain type of behavior, which (i) results from repeated exposure to a substance, (ii) causes a persistent, recurring desire to engage that type of activity that is very difficult to control, (iii) will worsen without treatment, (iv) remains capable of being triggered long after periods of abstinence, and (v) results in harmful consequences.*

First, Volkow's account violates SC2 since her definitions and characterizations are restricted to substance addictions and their effects on the brain. Two paradigmatic examples are as follows:

...[addiction is] a conditioned response [to the exposure] to the drug and/or drug-related stimuli that activates [the striato-thalamo-orbitofrontal] circuit and results in the intense drive to get the drug (consciously perceived as craving) and compulsive self-administration of the drug (consciously perceived as loss of control).⁴³²

Addiction can be viewed as a pathology in how importance is attached to stimuli that predict drug availability and how the brain regulates (chooses) behavioral output in response to those stimuli.⁴³³

Indeed, the brain disease model of addiction that Volkow defends has trouble accounting for behavioral addictions because it tends to focus on how the chemical components of addictive substances interact with the reward system when ingested. In particular, these substances are said to

⁴³² Volkow & Fowler (2000, p. 323).

⁴³³ Kalivas & Volkow (2005, p. 1410).

“hijack” the reward system, mimicking ordinary reward responses and resulting in dysfunctioning. The point here is that Volkow’s account is too narrowly focused on substance addictions and so will, at the very least, have trouble making behavioral addictions a subtype of *addiction* on her view.

Second, this account violates SC3, SC5, SC6, and SC7 that require addiction be kept distinct from its causes, such as what brought a particular addiction about, and its effects, such as harmful consequences of using. We can see this in both quotes above (“results in the intense drive,” “how the brain regulates behavioral output”), but also in the following:

Among the most insidious characteristics of drug addiction is the recurring desire to take drugs even after many years of abstinence... [and] the compromised ability of addicts to suppress drug seeking in response to that desire even when confronted with seriously adverse consequences.⁴³⁴

Addiction: The most severe, chronic stage of substance use disorder, in which there is a substantial loss of self-control, as indicated by compulsive drug taking despite the desire to stop taking the drug.⁴³⁵

...long-term exposure to drugs is a necessary condition for the development of addiction.⁴³⁶

Volkow’s account often requires that actual addicted behavior take place, that the addiction result from a history of exposure to the relevant substance, and that addiction results in harmful consequences. The latter is seen both in her own characterizations and her adoption of the DSM-V definition of ‘substance use disorder’. The harm condition is also part of how she understands what it means for an addict to be compelled – *continued use despite harm*.

Third, Volkow straightforwardly adopts a disease model, and so the account violates SC9. In particular, she holds that addiction is a disease of the mesolimbic dopamine system in the brain.

Fourth, while Volkow is obviously sensitive to the plethora of neuroscientific data on addiction, her account does still seem to violate the empirical sensitivity criterion SC10. This is because, as Heyman and others have argued, it is unclear how Volkow’s account can accommodate

⁴³⁴ Kalivas & Volkow (2005, p. 1423).

⁴³⁵ Volkow et al. (2016, p. 364).

⁴³⁶ Volkow et al. (2016, p. 368).

some of the behavioral and epidemiological data on addiction.⁴³⁷ This is because her account rests on an implausible notion of compulsion that is inconsistent with the fact that addicts often exhibit some control, typically end up quitting without treatment, and so forth.

Fifth, her stance on the brain disease question and her focus on neurobiology results in her account violating SC11 requiring an account be able to unify or incorporate distinct disciplinary perspectives as far as possible.

Sixth, violating SC3 and SC5-7 make it clear that the account also violates SC13. Volkow does sometimes speak of sensitivities and conditioned responses of the relevant brain systems. However, she does not have a genuinely dispositionalist account until the historical condition, behavioral manifestations, and harmful consequences are eliminated from it. These may be common or typical features that are causally related to addiction, but they are not part of what addiction is.

5.3.9 Jerome Wakefield

As with Pickard's view, there is much to like about Wakefield's account of addiction. While he defends a version of the disease account against Lewis's criticisms of the brain disease model, he does so without making many of the mistakes that Volkow and other brain disease proponents often make.⁴³⁸ For instance, he does not conflate systematic brain changes with disease or require harmful consequences for an addiction to be present.⁴³⁹ Of note is Wakefield's adherence to SC12 concerning the proper level of analysis. Wakefield engages the neuroscientific arguments of Lewis while still attempting to understand the nature of addiction in a way that can mesh with different disciplinary

⁴³⁷ Gene Heyman and Marc Lewis have been very vocal in this respect (Heyman, 2009, 2013; M. Lewis, 2015).

⁴³⁸ Lewis' book criticizes the brain disease model and he and Wakefield have a lively exchange on this issue (M. Lewis, 2015, 2017; Wakefield, 2017a, 2017b). Wakefield also notes some of the common errors the brain disease model proponents make and that he wants to avoid (Wakefield, 2017a, p. 39).

⁴³⁹ Wakefield *does* require harmful consequences in order for the addiction to count as an *addictive disorder*, which on his view is just an addiction (a certain biological dysfunction) plus harm (Wakefield, 2017a). This stems from his *harmful dysfunction account* of disease/disorder (these are synonymous on his view), which is a hybrid account requiring that any disease/disorder is both a biological dysfunction and harmful (Wakefield, 1992, 2014).

perspectives.⁴⁴⁰ Specifically, he sees addiction as a disruption of a person's capacity for deliberation and choice, which involves desires and underlying brain mechanisms. It is worth quoting Wakefield at length here to get a proper understanding of his view of addiction (emphases added):

Forms of *disruption of the desire/deliberation/choice system* that are not biologically designed to be peremptory can also occur. Such processes can so severely constrain the system that it functions outside the bounds of what is plausibly hypothesized to be normal variation. In such instances, *reasonable trade-offs between desires cannot be seriously contemplated, important balancing considerations are disregarded or summarily dismissed, and one motive very disproportionately dominates the deliberative process*, thus a dysfunction may be inferrable [51, 52]. Clearly, this disruption is a matter of degree once it passes over some fuzzy threshold that distinguishes the depth of the processes involved in, say, transient weakness of will in having too much pie at dinner from opiate addiction that destroys a family. In principle, severity of addiction should be measured in terms of *the degree and durability of the constriction of the desire/deliberation/choice system...A pathologically narrowed space of desire/deliberation/choice processes* is not the same as no space at all.⁴⁴¹

To summarize, Wakefield understands addiction as follows:

addiction: *a constricting or narrowing of the desire/deliberation/choice process which exceeds the threshold of variation that is within the range of proper functioning of that process.*

Before identifying the criteria that Wakefield's account violates, I should note that Wakefield's view is similar to Pickard's in another important way. While there is much to like about it, the account is lacking in detail regarding the metaphysical nature of addiction. Hence, while Wakefield's view only violates a few criteria and is comparatively much better than others on offer, the dispositionalist account I defend here can help to add to the substance that is lacking in a view like Wakefield's.

Now, the first two criteria this account violates are SC3 and SC5 requiring that addiction be kept distinct from effects such as its behavioral manifestations. The violation here comes in the context of significant ambiguity (or perhaps equivocation) on Wakefield's part. We can see from the passage above that it is really quite unclear whether Wakefield is referring to a constriction or

⁴⁴⁰ This is why, for instance, Wakefield can remain neutral about which neurobiological model of *how addiction works* is correct, since this is just the underlying basis of the addiction. Indeed, he states that "addiction is almost certain to be revealed eventually as not just one natural kind of brain process but many etiological kinds in which different things are going wrong that yield similarly problematic final pathways of compulsive drug taking, just like other major DSM categories of disorder such as depression and schizophrenia [53]. One should expect multiple etiologies..." (Wakefield, 2017a, p. 43).

⁴⁴¹ Wakefield (2017a, pp. 42–43).

narrowing of the relevant processes themselves (as they unfold, so-to-speak) or instead of the capacity for such constriction processes. Much of the language he uses suggests the former, including using a case of an actual instance of a weak-willed action as a comparison to addiction. Hence, it is reasonable to treat the criteria as violated yet easily fixable as opposed to not violated at all.

Second, the account seems to violate SC9 because Wakefield’s account makes addiction into a disease. The disruption of the desire/deliberation/choice system or processes is “pathological” in his words. However, I say ‘seems to violate’ because it is again unclear if Wakefield takes the disruption to be *necessarily* or *essentially* dysfunctional. The first two statements in the above passage suggest that dysfunctional disruptions are only one type of disruption. Whether non-dysfunctional disruptions are another possible type (that would still count as addictions) is not immediately clear. However, his larger exchange with Lewis does indicate that he sees addiction as inherently a biological dysfunction that is only sometimes harmful (and so only sometimes a disease). Again, given the lack of clarity here, it is reasonable to treat the criterion as violated (and perhaps easily fixable).

Lastly, the violations of SC3 and SC5 indicate that the account does not clearly make addiction into a disposition. Hence, it violates SC13. Again, this may be easily fixable without changing much of the substance of Wakefield’s account. Until Wakefield takes up this task, though, equivocating or remaining ambiguous on this point keeps his account from being genuinely dispositionalist.

5.3.10 Summary: Extant Accounts are Impaired or Metaphysically Insufficient

Table 4 summarizes the violations of the substantive criteria by the views considered here:

		Substantive Criterion												
		1	2	3	4	5	6	7	8	9	10	11	12	13
Extant View	Ainslie			x		x		x	x			x		x
	Foddy & Savulescu	x		x		x					x	x	x	
	Heather	x		x		x	x		x		x	x	x	x
	Heyman			x		x	x	x	x			x	x	x
	Levy								x	x	x	x		

Lewis	x		x		x	x					x	x	x
Pickard						x							
Volkow		x	x		x	x	x		x	x	x		x
Wakefield			x		x				x				x

Table 4: *Violations of Substantive Criteria by Extant View*⁴⁴²
(x = criterion violated)

The general moral of the story here is that existing accounts of addiction tend to be either impaired or metaphysically insufficient. On the one hand, they often violate one or more of the substantive criteria laid out here. Moreover, this says nothing of other complications these accounts may have, such as taking on problematic understandings of relevant concepts like *compulsion* or *disease*.⁴⁴³ On the other hand, existing accounts that violate one or very few of the criteria often lack a robust account of the metaphysics of addiction. For instance, while Pickard seems to explicitly take addiction to be a disposition, she does not tell us much about what this comes to metaphysically. Broad brush strokes in the right direction are certainly more helpful than painting the wrong picture, or no picture at all. However, if we are interested in really seeing the bigger picture for what it is, then we ought to pick up a fine-tooth comb at some point. Understanding addiction to be a disposition helps capture a number of the other criteria. But getting clearer about what this disposition is like can also help to fill in the details of accounts that are still scratching the surface, albeit the right one.

Let us now turn to evaluating existing ontological representations of addiction.

5.4 Evaluating Definitions in Existing Classification Systems and Vocabularies

In this section, I evaluate the definitions of ‘addiction’ in existing ontologies, as well as other classification systems and vocabularies that are relevant to the addiction literature, such as the

⁴⁴² SC4 is not violated. The reason this is included is because the ontological representations considered below are also judged by these criteria, and some violate SC4.

⁴⁴³ The brain disease model is notorious for taking on an implausible view of control/compulsion requiring literal irresistibility or something very close to it.

Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V) and the *National Institute on Drug Abuse (NIDA)* glossary. Here is the list of resources that I target in the below evaluations:

OBO Foundry Ontologies:

- (i) The Cell Line Ontology (CLO) (Sarntivijai et al., 2014);
- (ii) The ChEBI Integrated Role Ontology (CHIRO) (Hoyt et al., 2020);⁴⁴⁴
- (iii) The Experimental Factor Ontology (EFO) (Malone et al., 2010);
- (iv) The Human Disease Ontology (DOID) (Schriml et al., 2019);⁴⁴⁵
- (v) The Mental Disease Ontology (MFOMD), an extension of the Mental Functioning Ontology (MF) (Hastings et al., 2014; Hastings, le Novere, et al., 2012a; Hastings, Smith, et al., 2012);
- (vi) The Mammalian Phenotype Ontology (MP) (C. L. Smith & Eppig, 2009);
- (vii) The Neuro Behavior Ontology (NBO) (Gkoutos et al., 2012);
- (viii) The Obstetric and Neonatal Ontology (ONTONEO) (Barcellos Almeida & Farinelli, 2017);
- (ix) The Systemized Nomenclature of Medicine - Clinical Terms (SNOMED CT) (El-Sappagh et al., 2018); and,
- (x) The Unified Phenotype Ontology (uPHENO) (Matentzoglou et al., 2018).

Other Classification Systems and Vocabularies:

- (i) DSM-V (American Psychiatric Association, 2013);
- (ii) ICD-11 (World Health Organization, 2020);
- (iii) National Cancer Institute thesaurus (NCIt) (Fragoso et al., 2004);⁴⁴⁶
- (iv) National Institute of Drug Abuse (NIDA) glossary (NIDA, 2020); and,
- (v) U.S. Department of Health and Human Services (DHHS) glossary (OMHA, 2017).

These resources will be evaluated on the basis of both the substantive and methodological criteria.⁴⁴⁷ The former will be directed at the account of addiction itself – what the resource says

⁴⁴⁴ ChEBI is the Chemical Entities of Biological Interest Ontology, which does not itself have a class for addiction.

⁴⁴⁵ The DOID and the DSM-V did not use ‘addiction’ in the label, but only used the labels ‘substance-related disorder’ and ‘substance use disorder’, respectively (as opposed to other resources that used and defined both). I still consider the DOID because it is a highly successful and influential ontology, and terms like ‘substance use disorder’ and ‘substance-related disorder’ are often used interchangeably with ‘addiction’ in the literature. I still consider the DSM-V for similar reasons – it is highly influential and widely used, and the DSM-V criteria for *substance use disorder* are often used as either the definition of ‘addiction’ or the characterization of what is otherwise referred to as ‘addiction’.

⁴⁴⁶ The NCIt is developed with the intention of being compliant with the principles of the OBO Foundry. See their BioPortal homepage page where this intention is made explicit: <http://www.obofoundry.org/ontology/ncit.html>.

⁴⁴⁷ As above, I will use ‘SC’ and the relevant number, such as ‘SC1’, to refer to the substantive criterion violated. Similarly, for the methodological criteria I will use ‘MC’ and the relevant number, such as ‘MC1’.

addiction is – and the latter will be directed at the resource’s conformance (or lack thereof) to the relevant principles of best practice for ontology building.

5.4.1 Ontological Representations of Addiction

I start with ontologies that, as far as I can tell, explicitly attempt to represent addiction as an entity in its own right. This is in contrast to representing an entity that is a component of or merely related to the phenomenon of addiction, such as: **ABSTINENCE SYNDROME** (MFOMD, NCI), **BEHAVIORAL RESPONSE TO ADDICTIVE SUBSTANCE** (MP, NBO, uPHENO), **BIOBEHAVIORAL DETERMINANTS OF TOBACCO USE AND ADDICTION** (NCIt), **REGULATION OF ADDICTION RESPONSE** (NBO), or **SUSCEPTIBILITY TO TOBACCO ADDICTION** (DOID). Sometimes ‘addiction’ is used *as* the label, sometimes ‘addiction’ or one of its cognates is used *in* the label (as in ‘addictive behavior’), and sometimes ‘addiction’ or a cognate is explicitly specified as an *exact synonym* of the term used, as in this example from the EFO: **DRUG DEPENDENCE** *has_exact_synonym* **DRUG ADDICTION**.⁴⁴⁸ Importantly, in each case my contention is that the ontology developers took addiction *qua* entity to be the intended referent for the term.

Evaluations are mostly in alphabetical order according to the ontologies being evaluated. Sometimes multiple ontologies are grouped together because, for instance, they share the same definition or the same overarching problem (such as a missing definition). The evaluations will mostly be structured as follows. I will point to certain features of the ontology, such as a definition or type-subtype relation asserted, and then identify the criterion or criteria that feature violates. Moving from specific example to criteria violated will help the discussion flow more easily. It should also be noted that there are almost always more examples in the given ontology of the violation being considered.

⁴⁴⁸ See ‘drug dependence’ and synonyms in the EFO’s class hierarchy at the ontology lookup service here: <https://www.ebi.ac.uk/ols/ontologies/efo>.

Additionally, each ontology evaluated here is part of the Open Biological and Biomedical Ontology (OBO) Foundry’s suite of ontologies, and hence is explicitly committed to following the principles of best practice underlying BFO. Finally, these evaluations are restricted to the portions of the ontology representing addiction and closely related entities (such as the parent or sibling classes). Hence, there may be other portions of these ontologies not discussed here that properly conform to the principles of best practice, and others not discussed here that also violate the given criteria. A full evaluation of each ontology is beyond the scope of the present project.

5.4.1.1 CLO and SNOMED-CT: No Definition Provided

The Cell Line Ontology (CLO) and the Systematic Nomenclature of Medicine-Clinical Terms (SNOMED-CT) include the class **ADDICTION**. However, while the term is present, neither provides a definition. Thus, both CLO and SNOMED-CT already violate MC1 requiring that each term in the ontology be defined. Accordingly, both ontologies fail to meet MC1a-1c as well since they have no definition to adhere to these definition-related sub-criteria. Figure 17 below illustrates the lack of a definition in SNOMED-CT. Figure 18 below shows that the definition status ID for the SNOMED-CT class **ADDICTION** means *either* that it is not sufficiently defined by necessary conditions *or* that it is a primitive (the ID is used to mean both).⁴⁴⁹

Preferred Name	Addiction
Synonyms	Addiction (finding)
ID	http://purl.bioontology.org/ontology/SNOMEDCT/32709003
Active	1
altLabel	Addiction (finding)
DEFINITION STATUS ID	90000000000074008

Figure 17: No Definition for the SNOMED CT Class Addiction

⁴⁴⁹ The term is not used as a primitive since the class appears under at least five parent classes.

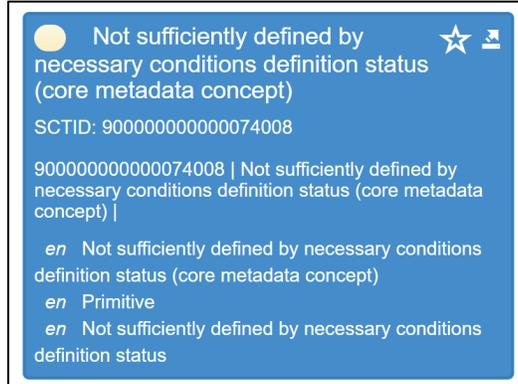


Figure 18: Meaning of Definition Status for Addiction in SNOMED CT⁴⁵⁰

Despite lacking a definition, we can still evaluate the ontologies’ classifications of addiction and some related kinds of entities, which can also provide some ground for evaluating their respective accounts of addiction. I start with CLO and then move to SNOMED-CT.

5.4.1.1.1 The Classification in CLO

Figure 19 below shows how CLO classifies **ADDICTIVE BEHAVIOR**, for which ‘addiction’ is given as an exact synonym.

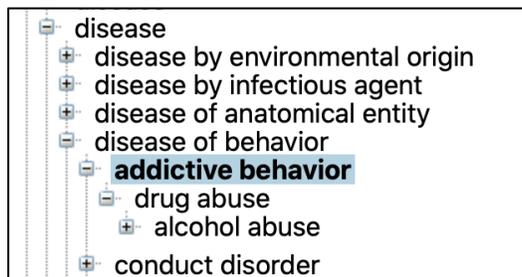


Figure 19: Classification of Addiction in CLO

First, the CLO class representing addiction is labelled ‘addictive behavior’, is classified as *subtype_of* **DISEASE OF BEHAVIOR**, and has **DRUG ABUSE** as a child class.

- (i) This violates SC3, SC5, and SC13 by identifying addiction with its behavioral manifestations, such as the addictive behavior an addict performs or the drug abuse that can cause an addiction. Indeed, ‘disease of behavior’ is defined in terms of “the manner that an organism conducts itself” and the “response to” stimulation or the environment.⁴⁵¹

⁴⁵⁰ See <https://browser.ihtsdotools.org/?perspective=full&conceptId1=900000000000074008&edition=MAIN/2020-03-09&release=&languages=en> for the source of the definition status.

⁴⁵¹ See the definitions of CLO classes here <https://biportal.bioontology.org/ontologies/CLO>.

- (ii) This also violates SC9 because it makes addiction a disease by definition.
- (iii) This also violates MC3b because the label confuses addiction with the behavior addicts typically perform (but which non-addicts can also perform) and so decreases clarity.

Second, CLO has two definitions of ‘disease’. One is from the Human Disease Ontology and is BFO-conformant: “a disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism.” However, the CLO class **ADDICTIVE BEHAVIOR** is a child of the class **DISEASE** that is connected to CLO’s other use of ‘disease’, defined as “a disposition that describes states of disease associated with a particular sample and/or organism.”

- (i) This violates MC2c because many children of each class labeled ‘disease’ are multiply inherited, appearing under both classes.
- (ii) This also seems to violate MC2a and MC2b because it jeopardizes the consistency of the *is_a* hierarchy by not adhering to the definition and transitivity of the *is_a* relation (MC2a and MC2b, respectively). If addiction is a behavior in CLO, as the label suggests, then it classifies a behavior (a BFO **PROCESS**) as a disposition (a BFO **SPECIFICALLY DEPENDENT CONTINUANT**).
- (iii) This also violates MC3b because the labels decrease clarity by using the same name for different entities.
- (iv) This also violates MC4 because the definition of ‘disease’ given for the CLO parent of **ADDICTIVE BEHAVIOR** (“a disposition that describes states...”) is not a realist definition. Neither diseases nor dispositions describe anything, and so the class is representing the *term* ‘disease’ and not the *entity in reality*.

Third, CLO **DISEASE** has **DISORDER** and several subtypes of **DISORDER** as children.

- (i) This violates either MC2 or MC5. If CLO **DISORDER** is understood in the BFO-conformant sense of the Human Disease Ontology, then CLO classifies a **MATERIAL ENTITY** as *subtype_of* **DISPOSITION** (an *is_a* inconsistency, MC2). If CLO **DISORDER** is not understood in this way, then CLO is not properly linkable back up to BFO or BFO-conformant ontologies (MC5).

5.4.1.1.2 The Classification in SNOMED-CT

Figure 20 below shows how SNOMED-CT classifies **ADDICTION**.⁴⁵²

⁴⁵² See the definitions of SNOMED-CT classes here <https://bioportal.bioontology.org/ontologies/SNOMEDCT>.

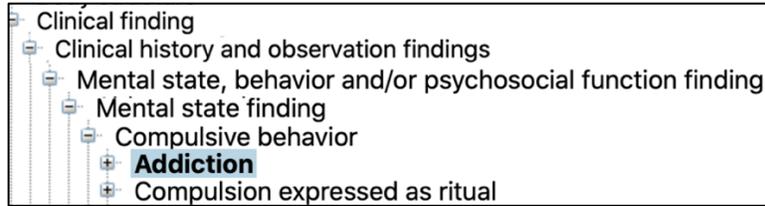


Figure 20: Classification of Addiction in SNOMED-CT

First, SNOMED-CT makes the class **ADDICTION** a *subtype_of* **COMPULSIVE BEHAVIOR**, which is a *subtype_of* **MENTAL STATE FINDING**.

- (i) This violates SC3, SC5, and SC13 because the classification confuses addiction for the behaviors that manifest it.
- (ii) This also violates SC10 because it makes addiction a kind of compulsion. This cannot easily accommodate empirical data showing that addicts can exercise some control.
- (iii) This also violates MC2a and MC2b related to the consistency of the *is_a* hierarchy. Behaviors are not mental states, and hence a compulsive behavior would not be a mental state finding.

Second, SNOMED-CT makes **ADDICTION** a *subtype_of* **CLINICAL FINDING**.

- (i) This violates MC4 and MC5 because this is a non-realist understanding of addiction, and hence it cannot properly link back up to BFO. A particular addiction might be an instance of a clinical finding. However, this is contingent and distinct from *what addiction is*. Addiction is not a finding; a *finding of addiction* is a finding. Hence, this classification is non-realist and relies on what we discover in clinical settings.

Third, SNOMED-CT has conjunctive and disjunctive classes such as **CLINICAL HISTORY AND OBSERVATION FINDINGS** and **MENTAL STATE, BEHAVIOR, AND/OR PSYCHOSOCIAL FUNCTION FINDING**.

- (i) This violates MC2d because it adds classes by conjunction and disjunction.
- (ii) This also violates MC3a and MC3c because it labels these classes with plural nouns and logical operators.

Fourth, SNOMED-CT has the class **PSYCHOACTIVE SUBSTANCE DEPENDENCE**, which is a *subtype_of* **PSYCHOACTIVE SUBSTANCE USE DISORDER**, and which has **ALCOHOLISM**, **ABSINTHE ADDICTION** and **SYNTHETIC CANNABINOID DEPENDENCE** as child classes.

- (i) This violates SC4 because it makes addiction – at least certain types, such as alcoholism and absinthe addiction – the same as dependence.

- (ii) This also seems to violate SC9 by making addiction a disease in the DSM-V sense.
- (iii) This also violates MC2a-2c because addiction is multiply inherited in SNOMED-CT as a child of both **COMPULSIVE BEHAVIOR** (and its parents) and **PSYCHOACTIVE SUBSTANCE DEPENDENCE** (and its parents). These are incompatible classes.
- (iv) This violates MC3b because the labeling here is quite confusing. Some addictions are dependence while addiction itself is a compulsive behavior, which is a mental state finding. It is hard to know how to make sense of this.

Fifth, SNOMED-CT has classes apparently representing some behavioral addictions, such as **GAMBLING DISORDER PREDOMINANTLY ONLINE**, and others apparently representing substance addictions, such as **ALCOHOLISM**, **ABSINTHE ADDICTION**, and **COMPULSIVE UNCONTROLLABLE DRUG TAKING**. These classes are not all subsumed under a common parent.

- (i) This violates SC2 because SNOMED-CT's classifications do not account for the fact that every substance and behavioral addiction will be *instance_of* **ADDICTION**.

5.4.1.2 CHIRO, MP, ONTONEO, and uPHENO: A Dependence with Certain Results

The ChEBI Integrated Role Ontology (CHIRO), the Mammalian Phenotype Ontology (MP), the Obstetric and Neonatal Ontology (ONTONEO), and the Unified Phenotype Ontology (uPHENO) all use the same definition of 'addiction', which is as follows:

addiction (MP:0002555) =def. *A strong dependence on an addictive compound such as alcohol or narcotics that results in uncontrollable cravings for such compounds.*⁴⁵³

According to these ontologies, addiction is just substance dependence that has particular kinds of effects. Consider first the criteria these ontologies violate with this definition.

- (i) This violates SC2 because it defines 'addiction' in terms of *substances* like alcohol and narcotics, and hence is not able to capture behavioral addictions.
- (ii) This also violates SC3 and SC13 because it conflates addiction with its effects by requiring that an addiction "results in uncontrollable cravings."
- (iii) This also violates SC4 because it equates addiction and dependence, or at least makes the former a kind of the latter.
- (iv) This also violates SC10 because it describes the required cravings as 'uncontrollable' and so cannot accommodate the evidence that addicts can sometimes exercise control.

⁴⁵³ See the definitions of MP classes here: <https://bioportal.bioontology.org/ontologies/MP>.

- (v) This also violates MC1a because the definition includes ‘addictive compound’, making the definition circular because ‘addiction’ is precisely what the definition is supposed to be explaining. ‘Addictive compound’ is not defined elsewhere and likely just means something like ‘substance that can cause an addiction’.
- (vi) This also violates MC1b because the definition introduces more complicated concepts that only raise further questions. Most importantly, these questions and concepts are not answered or defined elsewhere in the ontologies. For instance, we are left to our own devices about what ‘strong’ and ‘uncontrollable’ come to.

Consider next the problems with classification and the *is_a* hierarchy in these ontologies. First, MP is the source of the definition, and it classifies **ADDICTION** as shown in Figure 21 below.

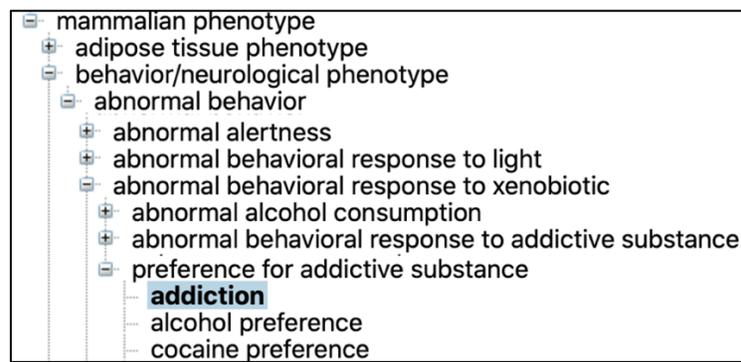


Figure 21: Classification of Addiction in MP

The classification here violates a number of methodological criteria regarding a consistent *is_a* hierarchy and linking back up to BFO.

- (i) This violates MC2a and MC2b because neither addiction nor dependence (as MP and the other ontologies define ‘addiction’) are preferences. Dependence is tolerance and the disposition to withdrawal. Moreover, many addicts do not *prefer* to continue using at all, since urge-like *wanting* comes apart from pleasure-associated *liking*.⁴⁵⁴ Moreover, we can see that **ADDICTION** is classified as a *subtype_of* **PREFERENCE** (likely a BFO **DISPOSITION**), **ABNORMAL BEHAVIOR** (a BFO **OCCURRENT**), and ultimately the MP root class **MAMMALIAN PHENOTYPE** (including both BFO **CONTINUANT** and BFO **OCCURRENT**).
- (ii) This also violates MC2d, then, because the root node **MAMMALIAN PHENOTYPE** combines continuants and occurrents through conjunction: any “observable morphological, physiological, behavioral and other characteristics of mammalian organisms that are manifested through development and lifespan.”
- (iii) This also violates MC4, requiring adherence to realism, since the definition of the MP root node entails that any entity in MP must be *observable*.

⁴⁵⁴ This is a well-documented and generally accepted finding (Berridge & Robinson, 2016; Robinson & Berridge, 1993).

- (iv) This also violates MC1c, requiring Aristotelian definitions. MP defines ‘addiction’ in terms of *dependence* but classifies it as *subtype_of* **PREFERENCE**. If **ADDICTION** *is_a* **DEPENDENCE**, then **DEPENDENCE** should be the parent. If the parent is **PREFERENCE**, then ‘addiction’ should be defined in terms of *preference*, not dependence.
- (v) This also violates MC5 because, given these problems, the ontologies cannot successfully link back up to BFO as they are currently constructed.

Finally, despite sharing the same definition, ONTONEO classifies **ADDICTION** differently than MP, and uPHENO has no classifications at all. The former is illustrated in Figure 22 below.

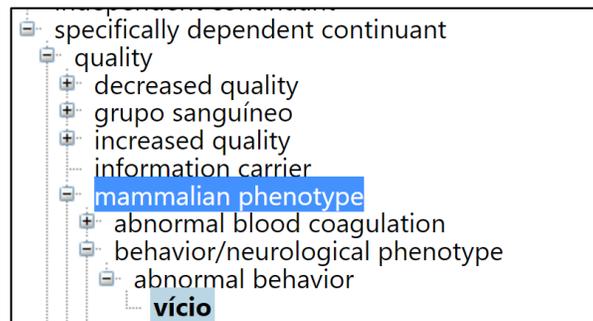


Figure 22: Classification of Addiction in ONTONEO⁴⁵⁵

Unlike MP and CHIRO, which have **MAMMALIAN PHENOTYPE** as the root node, ONTONEO classifies **MAMMALIAN PHENOTYPE** as a *subtype_of* BFO **QUALITY**, which is a **SPECIFICALLY DEPENDENT CONTINUANT**. This is inconsistent with MP’s definition of ‘mammalian phenotype’, which is supposed to allow both continuants and occurrents to count as mammalian phenotypes. uPHENO, on the other hand, uses MP’s definition of ‘addiction’ but has no classifications at all. Given the definition we saw above, ONTONEO’s classification exacerbates the problems with classification and thus further violates MC2 and its sub-criteria, as well as MC5 (linkability to BFO). uPHENO also violates MC2 and its sub-criteria since it lacks an *is_a* hierarchy.

5.4.1.3 DOID: Mental Disorder Involving Substance Abuse or Dependence

In the Human Disease Ontology (DOID), the best candidate for a classification of addiction is a child class of **SUBSTANCE-RELATED DISORDER**, such as **SUBSTANCE ABUSE** or **SUBSTANCE**

⁴⁵⁵ ONTONEO labels the class both ‘addiction’ and ‘vício’, with the latter being preferred. The latter is Portuguese for ‘addiction’, which was translated by the Brazilian editor of the definition, Fernanda Farinelli.

DEPENDENCE. Given that ‘addiction’ is missing from the label, let me say something about targeting the DOID for evaluation. First, consider the DOID definition of ‘substance-related disorder’:⁴⁵⁶

substance-related disorder (DOID:303) =def. *A disease of mental health involving the abuse or dependence on a substance that is ingested in order to produce a high, alter one’s senses, or otherwise affect functioning.*

While this class may not represent addiction itself, it seems clear that it would be the intended parent of **ADDICTION** within the DOID. Next, consider the definitions of two children of this class:

substance abuse (DOID:302) =def. *A substance-related disorder that involves a maladaptive pattern of substance use leading to significant impairment in functioning;*

substance dependence (DOID:9973) =def. *A substance-related disorder that involves the continued use of alcohol or other drugs despite problems related to use of the substance.*

Despite the use of Aristotelian definitions, it remains difficult to distinguish these two classes. However, the important point is that both are strikingly similar to the conception of addiction in the DSM-V (which we will come to below), and the DSM-V is widely considered to be an authority on identifying addiction. Hence, my evaluation focuses on whether **SUBSTANCE-RELATED DISORDER** would be a suitable candidate for a parent class of **ADDICTION**, as well as whether **SUBSTANCE ABUSE** or **SUBSTANCE DEPENDENCE** would be suitable candidates for representing addiction.

With this in mind, let us see which criteria would be violated under the assumptions that **ADDICTION** is meant to be either a child of or represented by one of the above classes.

- (i) This would violate SC2 because each class is substance-focused and so cannot capture behavioral addictions. Indeed, **PATHOLOGICAL GAMBLING** is classified as a subtype of **IMPULSE CONTROL DISORDER**, a *sibling* of **SUBSTANCE-RELATED DISORDER**.
- (ii) This would also violate SC3, SC5, and SC7, confusing addiction with its causes or behavioral manifestations. The first definition requires particular reasons for using the substance (a cause). The latter two require the addictive behavior be present (“pattern of substance use” and “involves the continued use”). This would also violate SC13 since it would make addiction a behavior instead of a disposition towards that behavior.
- (iii) The third definition would violate SC4, conflating addiction and dependence.
- (iv) The latter two definitions would violate SC6 because they require addiction be harmful.

⁴⁵⁶ See definitions of DOID classes here: <https://biportal.bioontology.org/ontologies/DOID>.

- (v) This would also violate SC9 since each makes addiction a disease by definition.
- (vi) This would also violate the unifying component of SC11 by focusing on *substance* addiction.
- (vii) This would also violate MC2a and MC2b regarding a consistent *is_a* hierarchy. This is because the DOID is conflating behaviors (a BFO **PROCESS**), disorders (a BFO **MATERIAL ENTITY**), and diseases (a BFO **DISPOSITION**). Substance abuse and dependence are characterized in terms of the behaviors that unfold, defined in terms of disorder, and ultimately classified as *subtype_of* **DISEASE**.
- (viii) This would also violate MC3b concerning clear and proper naming conventions.
- (ix) This would also violate MC5 due to many of the problems just listed, and in particular (vi) and (vii), because the ontology and its classes are not linkable to BFO.

5.4.1.4 EFO: State of Dependence or Disorders Related to Drug Exposure

The Experimental Factor Ontology (EFO) classifies addiction using the label ‘drug dependence’, which has ‘drug addiction’ (formerly the preferred label), ‘drug habituation’, and ‘drug use disorder’ as synonyms, among others. Hence, I make the plausible assumption that the EFO class **DRUG DEPENDENCE** is meant to represent addiction. The definition is as follows:⁴⁵⁷

drug dependence (EFO:0003809) =def. *A state, psychic and sometimes also physical, resulting from the interaction between a living organism and a drug, characterized by behavioral and other responses that always include a compulsion to take the drug on a continuous or periodic basis in order to experience its psychic effects, and sometimes to avoid the discomfort of its absence. Tolerance may or may not be present. A person may be dependent on more than one drug. Disorders related to drug abuse, the side effects of a medication, or toxin exposure.*

There is a lot going on in this definition but let us see which substantive criteria it violates.

- (i) This violates SC2 because it is substance-focused.⁴⁵⁸
- (ii) This also violates SC3 and SC7 because it conflates addiction with what causes or influences it. For instance, it requires that it “[result] from the interaction between” the addict and the drug and also that certain reasons influence its manifestation, such as “to experience its psychic effects” or “to avoid the discomfort of its absence.”
- (iii) This also violates SC4 by equating it with or making it a subtype of dependence.
- (iv) This also violates SC9 since **DRUG DEPENDENCE** is *subtype_of* **DISEASE** in EFO.
- (v) This also violates the unifying component of SC11 by focusing on substance addictions.

⁴⁵⁷ See definitions of the EFO classes here: <https://bioportal.bioontology.org/ontologies/EFO>.

⁴⁵⁸ **ADDICTIVE BEHAVIOUR** is also a class in EFO. However, this is almost certainly not meant to represent addiction since it is defined in terms of the observable activity of an organism that indicates an addiction is present. Moreover, neither ‘addiction’ nor ‘drug addiction’ are synonyms like they are for **DRUG DEPENDENCE**.

Now, EFO's *classification* of **DRUG DEPENDENCE** (and thus addiction) is also problematic.

Multiple inheritance occurs (see Figure 23 below), but the main chain of *is_a* relations is as follows:

DRUG DEPENDENCE *is_a* **DRUG-INDUCED MENTAL DISORDER** *is_a* **MENTAL OR BEHAVIOURAL DISORDER** *is_a* **PSYCHIATRIC DISORDER** *is_a* **DISEASE** *is_a* **DISPOSITION** *is_a* **MATERIAL PROPERTY** *is_a* **EXPERIMENTAL FACTOR**.

An illustration of the multiple inheritance in EFO is shown in Figure 23 below, which also serves to illustrate some of the violations of other methodological criteria listed below.

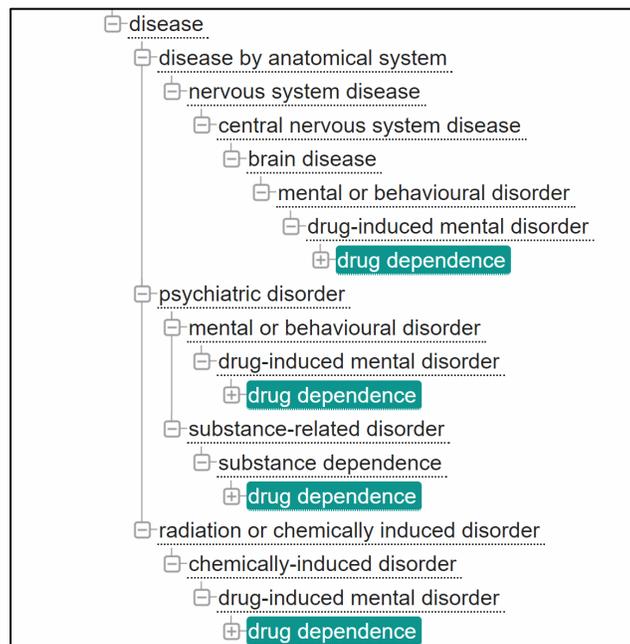


Figure 23: Classification of Drug Dependence in EFO

Here, then, are the methodological problems based on the above definition, the main chain of *is_a* relations listed above for **DRUG DEPENDENCE** in EFO, and Figure 23.

- (i) This violates MC1b regarding definitional clarity. For instance, there are two inconsistent definitions within the definition of 'drug dependence': the first sentence (roughly, the state of an organism of being disposed to compulsive behaviors toward a drug) and the last sentence (any disorder related to drug abuse, medication side effects, or toxin exposure). Moreover, 'compulsion' is required but is not defined in the ontology.
- (ii) This also violates MC2a and MC2b regarding the consistency of the *is_a* hierarchy because, like some ontologies above, EFO conflates **DISEASE** (a BFO **DISPOSITION**) and **DISORDER** (a BFO **MATERIAL ENTITY**).
- (iii) This also violates MC2c because multiple inheritance occurs with **DRUG DEPENDENCE**.

- (iv) This also violates MC2d because EFO asserts numerous classes through conjunction or disjunction of other classes. For instance, two parent classes of **DRUG DEPENDENCE** are **MENTAL OR BEHAVIOURAL DISORDER** and **RADIATION OR CHEMICALLY INDUCED DISORDER**.
- (v) This also violates MC3c because it labels its classes using logical operators.
- (vi) This also violates MC4 because the root node **EXPERIMENTAL FACTOR** is defined as “the variable aspects of an experiment design which can be used to describe an experiment, or set of experiments, in an increasingly detailed manner.” Entities like addiction do not describe anything. This belies a non-realist approach underlying EFO.
- (vii) This also violates MC5 by making EFO unable to successfully link back up to BFO.

5.4.1.5 MFOMD: A Harmful Mental Disease (or Disorder? or Disease Course?)

The Mental Disease Ontology (MFOMD) extends the Mental Functioning Ontology (MF), and defines ‘addiction disorder’ (what Janna Hastings and her coauthors simply labeled ‘addiction’ in their original paper) as follows:⁴⁵⁹

addiction disorder (MFOMD:0000046) =def. *A mental disease in which a person persists in the use of a mood-altering substance or in a behavior despite adverse consequences.*

Given this definition, let us first turn to the substantive criteria violated by MFOMD.

- (i) This violates SC3, SC5, and SC13 because it requires addicts “persist in the use of [some] substance or in a behavior,” which conflates addiction (the disposition) with its behavioral manifestations.
- (ii) This also violates SC6 because it requires harmful (“adverse”) consequences that occur alongside the persistent behavior of the addict.
- (iii) This also violates SC9 because it makes addiction a disease by definition.
- (iv) This also violates SC12 because neither desires nor impaired control are required.

In addition to problems with the substantive criteria, MFOMD also violates a number of methodological criteria. Figure 24 below shows how **ADDICTION DISORDER** is classified in MFOMD, as well as how it classifies **MENTAL DISEASE**. As seen in the definition (and the top of the figure), the former is defined in terms of the latter.

⁴⁵⁹ Their paper discusses the addition of **ADDICTION DISORDER** to MFOMD (Hastings, le Novere, et al., 2012a). See definitions of the MFOMD classes here: <https://bioportal.bioontology.org/ontologies/MFOMD>.

There are, it seems, at least two ways to understand this class based on the definition. First, it might be intended to represent addiction itself. Given the mistakes we have seen above, this is not an implausible interpretation. Second, it might be intended to represent a manifestation of addiction – that is, a behavioral response that results from having an addiction. Either way, there are problems with the NBO definition and classification.

Here are the substantive criteria the NBO definition violates (I use “*” to indicate where the violation only applies to the first, stronger interpretation).

- (i) This violates SC2 since it is substance-focused and cannot capture behavioral addiction.
- (ii) This also violates SC3*, SC5*, and SC7* because it confuses addiction for its behavioral manifestations, and it requires that it result from having developed a dependence.
- (iii) This also violates SC4 because, on either reading, it confuses addiction with dependence. Either way, the definition implies that dependence is necessarily involved.
- (iv) This also violates the unifying component of SC11 because, while it requires impaired control, it requires that addiction necessarily manifest in *uncontrollable* cravings. This is akin to the *literal irresistibility* conception of compulsion seen in the brain disease model.

Figure 25 below shows the classification of **ADDICTION RESPONSE** in NBO. The violations of the methodological criteria are independent of which reading of the definition is correct.

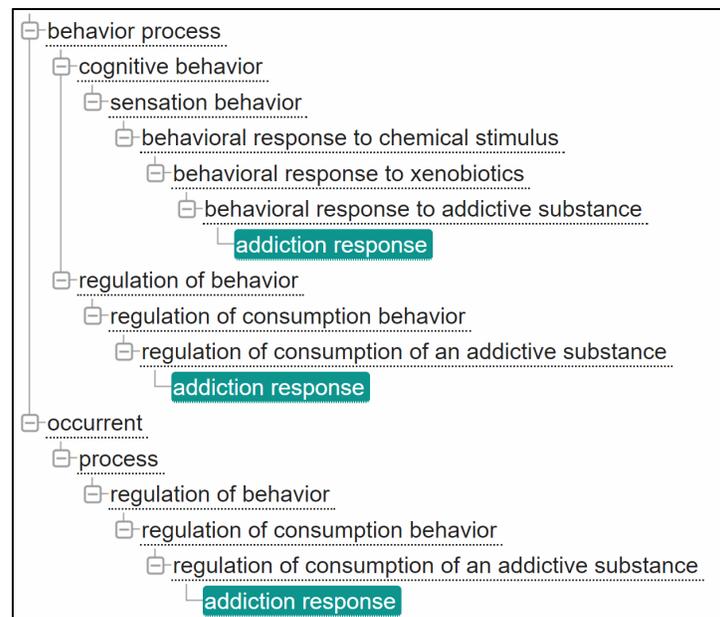


Figure 25: Classification of Addiction Response in NBO

Given the definition and Figure 25 above, the NBO violates the following methodological criteria:

- (i) This violates MC1a because, like the definition in MP above, the NBO definition is circular. It includes ‘addictive substance’, which likely means something like *a substance that can cause an addiction* and is therefore unhelpful at best and circular at worst.
- (ii) This also violates MC1b because ‘uncontrollable’ is used in the definition, but it is not made clear what this comes to or defined elsewhere in the ontology.
- (iii) This also violates MC2a and MC2b due to inconsistencies in the *is_a* hierarchy. NBO classifies **ADDICTION** as *subtype_of* both **BEHAVIOR** and **REGULATION OF BEHAVIOR**, yet something cannot be both a behavior and a regulation of that behavior.
- (iv) This also violates MC2c due to the multiple inheritance of **ADDICTION RESPONSE**.
- (v) This also violates MC3b (clear labeling) since ‘addiction response’ is as ambiguous as the definition that is meant to unpack it. It could imply either that the behavioral response is itself the addiction or that it is the typical outcome of an addiction.
- (vi) This also violates MC5 because, due to the foregoing, NBO cannot properly link up to BFO. For instance, Figure 25 above shows that NBO **BEHAVIOR PROCESS**, a parent of **ADDICTION RESPONSE**, is a *sibling* of BFO **OCCURRENT** rather than a *subclass* of it.

This concludes the evaluation of extant ontological representations of addiction. Before moving to the concluding sections of the chapter, I will consider some non-ontology classification systems and vocabularies that are important to addiction research.

5.4.2 Non-Ontology Classification Systems and Vocabularies on ‘Addiction’

I have canvassed a number of ontologies that attempt to represent addiction in one way or another. Here I look at other classification systems and vocabularies, which are not strictly speaking ontologies, and evaluate their definitions of ‘addiction’. Let me briefly say why.

First, this is partly in keeping with an adherence to the principles of best practice for ontology building, and in particular, the principle of *reuse* which helps to avoid reinventing the wheel. Second, with the exception of the National Cancer Institute Thesaurus, OBO Edition (NCIt), the definitions surveyed here are from classification systems and institutions that are widely used and involved in addiction research.⁴⁶² Hence, their relevance to the addiction literature makes them suitable candidates

⁴⁶² The DSM-V (American Psychiatric Association, 2013) and ICD-11 (World Health Organization, 2020) are gold standards in experimental and clinical contexts for addiction research. I also evaluate vocabularies of the *National Institute*

for evaluating their understandings of addiction. Lastly, the NCI is considered because, though it is somewhere in between an ontology and an ordinary dictionary, its developers explicitly state their intention to conform to the principles and standards of the OBO Foundry.⁴⁶³

5.4.2.1 The DSM-V: A Cluster of Criteria

I start with the representation of addiction in the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (DSM-V). The DSM-V does have some virtues. It can be a useful tool for helping clinicians determine whether, to what extent, and how an individual may need to undergo treatment for their purportedly addictive behaviors. Its developers are also explicitly committed to ontology-friendly principles, such as the *open world assumption* holding that our knowledge of the world is always revisable, a commitment to interoperability (such as with the widely used ICD-11), and the commitment to having an explicitly shared vocabulary for clinicians.⁴⁶⁴ Despite these virtues, the representation of addiction in the DSM-V suffers from a number of problems.

However, I should first note that there are at least two ways to understand the DSM-V's representation of addiction. On the one hand, the DSM-V might be offering a straight-forward *definition* of 'addiction', just as an ontology or vocabulary would do. That is, it might be trying to tell us *what addiction is*. Indeed, in their introduction to the chapter on *substance use and addictive disorders*, the editors call the clusters of diagnostic criteria they provide "the essential feature of a substance use disorder."⁴⁶⁵ On the other hand, the DSM-V might merely be offering a *characterization* of addiction, more closely resembling an entry in an encyclopedia than a dictionary or ontology. That is, it might

on *Drug Abuse* (NIDA) (NIDA, 2020) and the U.S. *Department of Health and Human Services* (DHHS) (OMHA, 2017). NIDA is a branch of the *National Institutes of Health* (NIH), which, along with organizations like the FDA and the CDC, is in turn funded by the DHHS. NIDA itself is responsible for upwards of 85% of funded addiction research (FNIDA, 2017; Muth, 2001; National Institutes of Health, 2005). Moreover, DHHS funding to organizations like the NIH and the *Substance Abuse and Mental Health Services Administration* in the fight against just *opioid* addiction exceeded \$4.2 billion in 2019 (OMHA, 2020). Hence, all of these resources are well worth consideration.

⁴⁶³ See their BioPortal homepage page where this is made explicit: <http://www.obofoundry.org/ontology/ncit.html>.

⁴⁶⁴ See the DSM-V (American Psychiatric Association, 2013, pp. 5–12).

⁴⁶⁵ See the DSM-V (American Psychiatric Association, 2013, p. 483).

only be trying to tell us *about addiction*, such as its typical signs and characteristic effects. The introductory chapter offers guidance on the manual's construction, organization, and use, but it is actually quite ambiguous about this point. Consider the following passages that come down on opposite sides of this *definition vs. characterization* issue:

Definition-friendly:

The *individual disorder definitions that constitute the operationalized sets of diagnostic criteria* provide the core of DSM-5 for clinical and research purposes. These criteria have been subjected to scientific review, albeit to varying degrees, and many disorders have undergone field testing for interrater reliability.⁴⁶⁶

Definition-unfriendly:

The symptoms contained in the respective diagnostic *criteria sets do not constitute comprehensive definitions of underlying disorders*, which encompass cognitive, emotional, behavioral, and physiological processes that are far more complex than can be described in these brief summaries. Rather, *they are intended to summarize characteristic syndromes of signs and symptoms that point to an underlying disorder* with a characteristic developmental history, biological and environmental risk factors, neuropsychological and physiological correlates, and typical clinical course.⁴⁶⁷

Two inconsistent conclusions follow. First, there are definitions of the particular disorders found in the DSM-V, and these definitions constitute the criteria for each respective disorder. Second, each disorder's criteria also *do not* constitute their definitions but are merely indications of the possibility that it is present. I exploit this ambiguity and evaluate addiction in the DSM-V under the assumption that it provides a definition of 'addiction'. First, it is a useful exercise to see how one of the core documents in the addiction literature would fare at *representing* and *classifying* the phenomenon we call 'addiction'. This is especially salient because the document is constantly used throughout the U.S. to diagnose people with – that is, to *identify* – substance use disorder (addiction). Second, suppose the DSM-V criteria are only meant to characterize addiction by way of common (though non-defining) features. Even still, this only strengthens my claim that there is a need for a better, more comprehensive, and more philosophically rigorous representation (or ontology) of addiction.

⁴⁶⁶ See the DSM-V (American Psychiatric Association, 2013, p. 10). Emphases are added to help illustrate the point.

⁴⁶⁷ See the DSM-V (American Psychiatric Association, 2013, p. 19). Emphases are added to help illustrate the point.

Hence, I focus on the eleven criteria for each particular *addictive disorder*, which the DSM-V separates into four groups: *impaired control*, *social impairment*, *risky use*, and *pharmacological criteria*.⁴⁶⁸ These can be found in Table 5 below using the case of *alcohol use disorder* as the example.⁴⁶⁹

CATEGORY	DSM-V CRITERION	SUMMARY
Impaired Control	1. Alcohol is often taken in larger amounts or over a longer period than was intended.	Use more than intended
	2. There is a persistent desire or unsuccessful efforts to cut down or control alcohol use.	Persistent desire to control use and/or failed attempts to do so
	3. A great deal of time is spent in activities necessary to obtain alcohol, use alcohol, or recover from its effects.	Time dominated by use, pursuit, and/or recovery
	4. Craving, or a strong desire or urge to use alcohol.	Craving/Desire
Social Impairment	5. Recurrent alcohol use resulting in a failure to fulfill major role obligations at work, school, or home.	Disruption of social obligations
	6. Continued alcohol use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of alcohol.	Continued use despite social problems
	7. Important social, occupational, or recreational activities are given up or reduced because of alcohol use.	Thwarting of social activities
Risky Use	8. Recurrent alcohol use in situations in which it is physically hazardous.	Use in harmful situations
	9. Alcohol use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by alcohol.	Use despite knowledge of existing harmful effects
Pharmacological Criteria	10. Tolerance, as defined by either: a. A need for markedly increased amounts of alcohol to achieve intoxication or desired effect. b. A markedly diminished effect with continued use of the same amount of alcohol.	Need to use more to get the same psychoactive effects
	11. Withdrawal, as manifested by either of: a. The characteristic withdrawal syndrome for alcohol. b. Alcohol is taken to relieve or avoid withdrawal symptoms.	Get sick when use is stopped or using in order to avoid such sickness

Table 5: Summary of DSM-V Criteria for Alcohol Use Disorder⁴⁷⁰

⁴⁶⁸ See the DSM-V for discussion of these divisions (American Psychiatric Association, 2013, p. 483).

⁴⁶⁹ Criteria are offered for each individual addictive disorder. However, the criteria for each listed *substance use disorder* are roughly the same. Each just substitutes the respective substance (alcohol, opioids, tobacco) in the criteria. Moreover, while the criteria for *gambling disorder* are slightly different (for instance, eliminating the withdrawal criteria), they very closely mimic the criteria for the substance-based addictive disorders, including criteria for impaired control, risky use, and social impairment. See the DSM-V for the full list of *gambling disorder* criteria (American Psychiatric Association, 2013, p. 585).

⁴⁷⁰ The DSM-V specifies three possible degrees of addiction: *mild*, *moderate*, and *severe* (American Psychiatric Association, 2013, p. 493). A mild case of, for instance, *alcohol use disorder*, requires the presence of any 2-3 of the criteria; a moderate

With the DSM-V criteria in full view, let us turn to the violations of our substantive and methodological criteria from Section 5.2. Consider first the violations of the substantive criteria.

- (i) The DSM-V violates SC3, SC5, and SC6, all of which have to do with conflating addiction with different kinds of effects or manifestations of addiction. This is because it is heavily focused on behavioral manifestations of addiction, often very specific ones like engaging in the addicted behavior in risky circumstances.
- (ii) The DSM-V also violates SC4 because it allows dependence (the combination of its two pharmacological criteria) to count as (a mild) addiction.
- (iii) The DSM-V's second criterion violates SC8 since it is inconsistent with willful addicts.
- (iv) The DSM-V also violates SC9 because it counts addiction as a disease/disorder by definition (simply in virtue of being in the manual).
- (v) The DSM-V also violates the impaired control condition of SC12 since no criterion listed is sufficient for impaired control. Like the problems in (i) above, they are merely possible indicators of impaired control.
- (vi) The DSM-V also violates SC13 because it fails to make addiction a disposition (mostly due to the problems mentioned in (i) directly above).

Consider next the violations of the methodological criteria.

- (i) The DSM-V violates MC1b and MC1c because, respectively, the definition impairs clarity and is not in Aristotelian form. The former is true because, given the degree schema used, two mild or moderate addictions can fail to have any features in common, and even two severe addictions would only need to have one feature in common.
- (ii) The DSM-V also violates MC2 because there is no real *is_a* hierarchy to be evaluated. At best, we can infer that **ADDICTION is_a MENTAL DISORDER** because it appears in the DSM-V in the first place. Still, it provides no definition for the latter. Moreover, the criteria violate MC2d in particular because the cluster methodology makes **ADDICTION** into a disjunctively defined class.
- (iii) The DSM-V also violates MC5 since it is not linkable to BFO.

5.4.2.2 The ICD-11: Classifying Addiction

The *International Classification of Diseases, 11th Edition* (ICD-11) is described by its developers as “the international standard diagnostic classification for all general epidemiological and many health management purposes...[which has evolved to become] a comprehensive classification system for use

case requires the presence of any 4-5 criteria; and a severe case requires the presence of any 6 or more of the criteria. Furthermore, the criteria need to have been present within the last 12-months prior to diagnosis.

in mortality, morbidity, casemix, quality measurement and patient safety.”⁴⁷¹ Before turning to the evaluation, a preliminary point is in order.

As with the DSM-V, one might object that the ICD-11 is not intended to offer *definitions*, but only *characterizations* for certain practical purposes like keeping statistical records. However, consider again the utility of evaluating such an important document to the addiction literature as if it were telling us *what addiction is*. Additionally, the ICD-11 is highly important and influential in addiction research, and mental health research more broadly. As the developers tell us, its scope of importance is wider than the DSM-V, being an *international* standard that helps dictate where most *global* health care resources end up. What is more, the developers’ preferred companion terminology is SNOMED-CT, and we already saw above that this ontology currently has no definition of ‘addiction’.⁴⁷² Thus, it is a worthwhile exercise to determine how well this important document represents addiction.⁴⁷³

Now, let us turn to the ICD-11 representation of addiction to be evaluated.⁴⁷⁴ As with the DSM-V evaluation, I use the particular ICD-11 class **ALCOHOL DEPENDENCE** as the example.⁴⁷⁵ Here, then, is the relevant ICD-11 entry:

*Alcohol dependence is a disorder of regulation of alcohol use arising from repeated or continuous use of alcohol. The characteristic feature is a strong internal drive to use alcohol, which is manifested by impaired ability to control use, increasing priority given to use over other activities and persistence of use despite harm or negative consequences.*⁴⁷⁶

⁴⁷¹ This is from the *ICD-11 Reference Guide* (World Health Organization, 2020, Sect. 1.1), available here: <https://icd.who.int/icd11refguide/en/index.html>. The *ICD-11* is available here: <https://icd.who.int/browse11/l-m/en>.

⁴⁷² The connection to SNOMED-CT is stated explicitly in the ICD-11 (World Health Organization, 2020, Sect. 1.1).

⁴⁷³ Indeed, given the scope and use of the DSM-V and the ICD-11, it is hard to imagine what work is left for any other resource in identifying cases of addiction, and providing us with information on its rates of prevalence, mortality, and the like. Importantly, though, how could such work be done without knowing what addiction is? At some point, the “we’re not actually defining it” response simply becomes unacceptable.

⁴⁷⁴ I focus on the ICD-11 classes most intuitively representative of addiction: those falling under the parent class **DISORDERS DUE TO SUBSTANCE USE OR ADDICTIVE BEHAVIOURS** (where else would addiction be?). The ICD-11 uses the label ‘dependence’ for these classes. Given that the entry makes reference to “physiological features of dependence” as distinct from the other features listed, this class is not intended to represent mere physical dependence.

⁴⁷⁵ As with the DSM-V, the ICD-11 separates descriptions and criteria for each type of substance addiction under different classes (such as **ALCOHOL DEPENDENCE** and **CANNABIS ADDICTION**), but each is roughly the same, substituting the name of each substance in the description/criteria. Also, as with the DSM-V, gambling disorder is slightly different.

⁴⁷⁶ See the ICD-11 (World Health Organization, 2020, Sect. 6C40.2). Here is the rest of the entry, which is more accurately understood as added information rather than definitional content: “These experiences are often accompanied by a subjective sensation of urge or craving to use alcohol. Physiological features of dependence may also be present, including tolerance to the effects of alcohol, withdrawal symptoms following cessation or reduction in use of alcohol, or repeated use of alcohol or pharmacologically similar substances to prevent or alleviate withdrawal symptoms. The features of

Consider first the substantive criteria that this account violates.

- (i) This violates SC3, SC5, and SC7. First, it confuses addiction with its causes by making an historical condition (specifically, repeated use) necessary to what an addiction is. Second, following the National Institute on Drug Abuse (discussed below) and the DSM-V, it makes *continued use despite harmful consequences* essential to addiction.
- (ii) This also violates SC6 because the account includes a harm condition.
- (iii) This also violates SC9 because it makes addiction a disease by definition. As with the DSM-V, this comes along simply by being included in the classification system (it is the International Classification of *Diseases*).
- (iv) This also violates SC13 because of the above problems and because we do not know what the ICD-11 means by ‘disorder’. Hence, it is reasonable to conclude that this account does not clearly see addiction as a disposition.

Turn next to the methodological criteria that the ICD-11 violates. Figure 26 below shows the type-subtype relations going from the ICD-11 class **ALCOHOL DEPENDENCE** to one of the many root nodes in the ICD-11: **MENTAL, BEHAVIORAL, OR NEURODEVELOPMENTAL DISORDERS**.

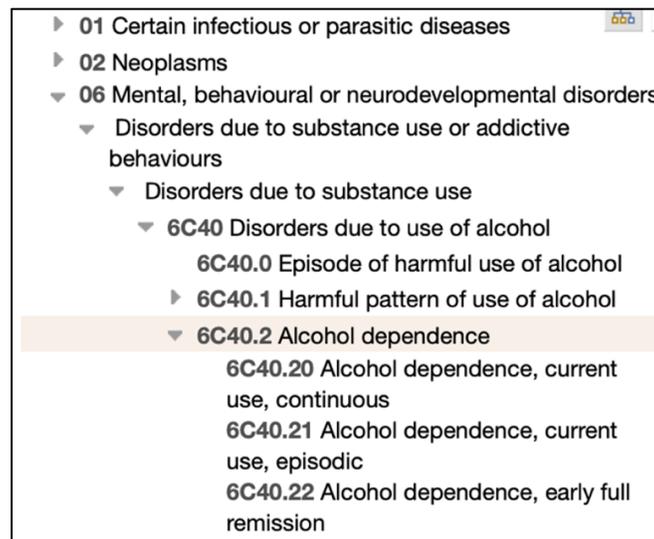


Figure 26: Classification of Alcohol Dependence in ICD-11

Using both the definition and the classification structure, here are the methodological criteria violated.

- (i) This violates MC1b because the definition invokes more complicated concepts that it does not define – ‘addiction’ is defined in terms of *disorder*, but this is not unpacked.

dependence are usually evident over a period of at least 12 months, but the diagnosis may be made if alcohol use is continuous (daily or almost daily) for at least 1 month.”

- (ii) This also violates MC2a and MC2b because of the inconsistency in the *is_a* hierarchy. The ICD-11 conflates the two fundamental categories from BFO, **CONTINUANTS** and **OCCURRENTS**, because it allows processes like **HARMFUL PATTERN OF USE OF ALCOHOL** to be a *subtype_of* **DISORDER**, which is a BFO **MATERIAL ENTITY**.⁴⁷⁷
- (iii) This also violates MC2d because the ICD-11 has classes asserted through disjunction, such as **MENTAL, BEHAVIOURAL, OR NEURODEVELOPMENTAL DISORDERS** and **DISORDERS DUE TO SUBSTANCE USE OR ADDICTIVE BEHAVIOURS**.
- (iv) This also violates MC3a-c because the proper naming conventions are violated. The violations of MC3a and MC3c can clearly be seen in the use of plural nouns and logical operators in the labels. An example of the violation of MC3b, regarding clarity, is the use of ‘dependence’ in the label when this does not refer to dependence as it is normally used (tolerance plus disposition to withdrawal).
- (v) This also violates MC4 concerning realism. While it is not visible in the figure above, the ICD-11 has the class **ALCOHOL DEPENDENCE, UNSPECIFIED** as a sibling of **ALCOHOL DEPENDENCE** (and the same for each kind of substance addiction). This belies a conception of the entities represented in the ICD-11 as dependent on what we know about them, as opposed to how they are in reality independent of our knowledge.
- (vi) This also violates MC5 because it is not linkable to BFO. This is partly due to both the classification problems above and the fact that ICD-11’s root nodes do not straightforwardly link to a class in BFO or a BFO-compliant ontology.⁴⁷⁸

In my view, the ICD-11, like the DSM-V, suffers from a kind of dilemma. On the one hand, we might interpret these documents as attempting to define ‘addiction’ – or attempting to directly represent it – because of the language in the documents and the way they are standardly used. If this is right, then a number of problems are present, both within the definitions themselves and with the broader classification systems of which they are a part. These are the problems I have outlined above.

On the other hand, we might interpret these documents as merely *characterizing* addiction or its typical features and effects. If so, then at least two other problems remain. First, this seems inconsistent with the way these documents are used. How can the DSM-V, for instance, be used to *diagnose* someone as an addict – that is, a person with an addiction – if it offers no account of what addiction is? An objector might persist that it offers an account of *evidence for* addiction, and so justifies

⁴⁷⁷ Since ICD-11 does not define ‘disorder’, I use the BFO understanding as *the material basis of a disease*.

⁴⁷⁸ Keep in mind that by ‘not linkable’ I only mean that it is not linkable as it is. Of course, many bad ontologies or classification systems *can* link to BFO once the appropriate modifications are made.

diagnoses on this basis. To that, I would reply with a question: “What is evidence for addiction when ‘addiction’ is undefined?” On this *evidence only* view, a DSM-V diagnosis of *substance use disorder* can only amount to a diagnosis of the specific criteria that are detected (withdrawal, tolerance, using in risky scenarios, and so on), and not something else, addiction, which supposedly underlies the outward signs. The latter requires an account of what addiction is.

Second, if these documents offer no attempt to define ‘addiction’ or to represent addiction directly, then there is a significant gap in the literature, owing partly to the problems with existing ontologies that include addiction. The DSM-V and ICD-11 are gold standards for classifying, diagnosing, and studying medical disorders (for which they include addiction). If they do not tell us what addiction is, then something else ought to. I have shown why existing ontologies’ attempts to do this are problematic. Thus, this horn of the dilemma leaves the DSM-V and the ICD-11 no better off.

I turn now to definitions of ‘addiction’ from three medical glossaries.

5.4.2.3 Medical Vocabularies: The NCI, NIDA, and the DHHS

I begin with the National Cancer Institute Thesaurus, OBO Edition (NCIt). The NCIt is a glossary which takes the original NCI Thesaurus and “aims to increase [its] integration...with OBO Library ontologies.”⁴⁷⁹ I think this is the right goal. The NCIt developers do express that this OBO Edition “should be considered experimental.”⁴⁸⁰ Nonetheless, it remains worthwhile to examine their experimental definition of ‘addiction’ for the reasons discussed above. The NCIt offers two definitions, one being dubbed the ‘alternative definition’, and they are as follows:

addiction (NCIT:C3479) =def. *Physical and/or psychological dependence to any substance.*
(alternative definition: *Uncontrollable craving, seeking, and use of a substance such as a drug or alcohol.*)⁴⁸¹

⁴⁷⁹ The OBO Foundry (2020a). The NCIt homepage can be found here: <http://obofoundry.org/ontology/ncit.html>.

⁴⁸⁰ The OBO Foundry (2020a).

⁴⁸¹ See definitions of the NCIt classes here: <https://bioportal.bioontology.org/ontologies/NCIT>.

These definitions violate both substantive and methodological criteria. I start with the former.

- (i) This violates SC2 because it is solely focused on substance addictions.
- (ii) This also violates SC3, SC5, and SC13 because it conflates addiction with its behavioral manifestations, claiming addiction requires actual “craving, seeking, and use of a substance” as opposed to a certain disposition towards such things.
- (iii) This also violates SC4 because it makes addiction into dependence.
- (iv) This also violates SC9 by making addiction a disease by definition (what it calls a ‘behavioral disorder’), as can be seen in Figure 27 below.
- (v) This also violates the unifying condition of SC11 because it invokes the notion that addicts necessarily undergo *uncontrollable cravings*, invoking the highly controversial *literal irresistibility* conception of compulsion taken up by the brain disease model.
- (vi) This also violates the empirical adequacy condition of SC10 due to (iii) and (v) above. Both ideas are hard to square with what we know about addiction (it comes apart from dependence and addicts can exhibit control).

Figure 27 below illustrates the classification of **ADDICTION** in NCI.

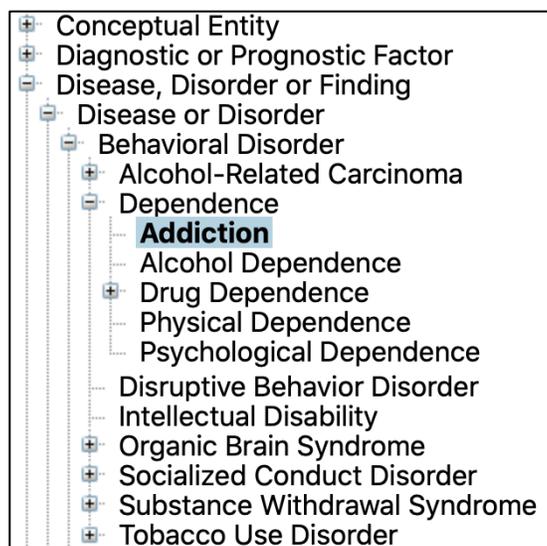


Figure 27: Classification of Addiction in NCI

Together with the above definitions, this leads to the following methodological criteria violations.

- (i) This violates MC1b regarding clear definitions using simpler terms than the terms being defined. The definition describes addiction as ‘uncontrollable’ without telling us what this means. There are also two distinct, inconsistent definitions, which diminishes clarity.
- (ii) This also violates MC2a, MC2b, and MC2d regarding a consistent *is_a* hierarchy because it conflates inconsistent entity types and asserts classes through conjunction and disjunction. For example, the class **DISEASE, DISORDER OR FINDING** is defined as “A condition that is relevant to human neoplasms and non-neoplastic disorders. This includes observations, test results, history and other concepts relevant to the

characterization of human pathologic conditions.” This mistakenly allows types of both **CONTINUANT** and **OCCURRENT** to share the same parent class.

- (iii) This also violates MC4 requiring a realist methodology. The definition in (ii) directly above makes certain entities reliant on whether we observe them and also appeals to concepts of entities instead of the entities themselves.
- (iv) This also violates MC3b and MC3c regarding clarity and the use of logical operators in labels. Regarding clarity, consider that **ADDICTION** is a *subtype_of* **DEPENDENCE**, is defined as “physical and/or psychological dependence to any substance,” and yet four different kinds of dependence are *siblings* of **ADDICTION**. These include **PHYSICAL DEPENDENCE** and **PSYCHOLOGICAL DEPENDENCE**, which is confusing at best.
- (v) This also violates MC5 because of the above problems, namely those involving the conflation of **CONTINUANT** and **OCCURRENT**. Hence, NCI cannot link up to BFO.

Consider next the glossaries from the National Institute on Drug Abuse (NIDA) and its parent organization, the U.S. Department of Health and Human Services (DHHS). Despite their tight connection, different definitions of ‘addiction’ are given. For instance, the NIDA definition requires harmful consequences while the DHHS definition does not. The DHHS definition also appeals to the DSM-V criteria, while the NIDA definition does not. Since NIDA is a branch of DHHS, I provide both definitions below and evaluate them together as was done with the two NCI definitions.

NIDA:

addiction =def. *A chronic, relapsing disorder characterized by compulsive (or difficult to control) drug seeking and use despite harmful consequences, as well as long-lasting changes in the brain.*

DHHS:

addiction =def. *The most severe form of substance use disorder, associated with compulsive or uncontrolled use of one or more substances. Addiction is a chronic brain disease that has the potential for both recurrence (relapse) and recovery.*

Consider first the violations of the substantive criteria.

- (i) These violate SC2 because they are solely focused on substance addictions.
- (ii) These also violate SC3, SC5, and SC13 because they conflate addiction with its behavioral manifestations (substance use, relapse, drug seeking) and effects (long-term brain changes) instead of making it a disposition towards these things.
- (iii) These also violate SC6 because they require harmful consequences. NIDA does this explicitly. DHHS does this because it invokes the DSM-V criteria in the definition, which requires that addiction be harmful.
- (iv) These also violate SC9 by making addiction a disease by definition.

- (v) These also violate the unifying component of SC11 because they focus heavily on the brain. DHHS does this more explicitly in the definition, while NIDA does it through the context within which the definition is given. Specifically, NIDA is the home of the brain disease model of addiction.

Consider next the violations of the methodological criteria.

- (i) These violate MC1b and MC1c related to the clarity and Aristotelian form of the definitions. They fail to go from complex to more simpler terms because they do not provide definitions for terms they invoke like ‘disease’, ‘disorder’, or ‘compulsion’.
- (ii) These also violate MC2 generally because there is no *is_a* structure being used at all.⁴⁸²
- (iii) The DHHS definition violates MC2a and MC2b if we assume a BFO-compliant understanding of disease and disorder, since it would then conflate a BFO **DISPOSITION** with a BFO **MATERIAL ENTITY**.
- (iv) The DHHS definition also violates MC2d by invoking the DSM-V criteria because it defines ‘addiction’ by asserting a disjunctively defined class.
- (v) Both also violate MC5 because they cannot successfully link up to BFO.

This concludes the evaluation of existing representations of addiction in extant ontologies and other classification systems and vocabularies important to the addiction literature.

5.4.3 Extant Resources Representing Addiction are Flawed

Table 6 on the following page summarizes the violations of the substantive and methodological criteria by the resources just considered. The moral of the story is the same as that from the evaluations of extant accounts: by and large, existing representations of addiction are flawed. Whether they are ontologies proper or other classification systems and vocabularies relevant to the addiction literature, they fail to meet many of the criteria laid out in Section 5.2. This is true of the methodological criteria, which apply in the sense that these resources serve as artifacts that attempt to represent addiction. Recall that this is precisely what well-built ontologies are meant to do.

⁴⁸² Of course, these are glossaries and not ontologies. However, since the evaluation is about how well current resources important to the addiction literature represent addiction, it is justifiable to impose the ontology principles of best practice on these resources because (realist) ontologies are for representing entities in the world. These resources are representing an entity in the world, addiction, and thus should be doing so according to the principles of best practice.

		Substantive Criteria													Methodological Criteria													
		1	2	3	4	5	6	7	8	9	10	11	12	13	1a	1b	1c	2a	2b	2c	2d	3a	3b	3c	4	5		
		Resource Representing Addiction																										
NIDA, DHHS		x	x		x	x			x		x			x		x	x	x	x	x						x		
NCIt			x	x	x	x				x	x	x		x			x	x	x							x	x	
ICD-11				x		x	x				x			x			x			x	x	x	x	x	x	x	x	
DSM-V				x	x	x		x	x			x		x			x	x	x	x							x	
NBO			x	x	x	x		x				x				x	x		x	x	x			x			x	
MFOMD					x		x					x	x					x	x					x			x	
EFO			x	x	x			x				x					x	x	x	x				x	x	x	x	
DOID			x	x	x	x	x		x					x				x	x					x			x	
CHIHO, ONTONEO, MP, uPHENO			x	x	x						x			x		x	x	x			x						x	x
SNOMED			x	x	x	x				x	x			x				x	x	x	x	x	x	x	x	x	x	
CLO				x		x					x			x		x	x	x	x				x			x	x	

Table 6: Violations of Substantive and Methodological Criteria by Resource
(x = criterion violated)

But this is also true of the substantive criteria, which apply in the sense that ontologies, classification systems, and vocabularies are providing an account of what addiction is in just the way an author's view does. Even though ontology-building in general is meant to be neutral with respect to which substantive view is right (come one, come all), it remains true that any particular term or class in an ontology will provide *some* particular view of that entity, however thin. For instance, an ontology might classify addiction as a disposition or as a process, two completely different kinds of entities. Once classified and then defined (ideally in terms of essential features), we have at least the bare bones of a substantive account of that entity. This is all I had in mind when evaluating the resources above according to the substantive criteria.

Here, then, is a restatement of the defense of premise (4) of my overarching argument:

- (4a) The dispositionalist account fares better than prominent extant competitor views with respect to satisfying the substantive desiderata; [Section 5.3]
- (4b) The dispositionalist account fares better than existing ontological representations of addiction with respect to satisfying the substantive and methodological desiderata; [Section 5.4]
- (4c) If (4a) and (4b), then the dispositionalist account of addiction fares better than competitor accounts, both philosophical and ontological; [trivial]
- (4) Hence, the dispositionalist account of addiction fares better than competitor accounts, both philosophical and ontological. [(4a)-(4c)]

5.4.4 AddictO to the Rescue?

Before turning to the review of where we have come thus far, I need to briefly comment on the recent project that Janna Hastings and Robert West have begun, developing an ontology of addiction dubbed 'AddictO'.⁴⁸³ First and most importantly, this project is relatively brand new and so the ontology is still very much under development. Second, I have very recently become involved in this project, and so will be engaged in the process of helping to develop this ontology in the right

⁴⁸³ See Hastings et al. (2020). The homepage for AddictO, where you can download the current version of the ontology or search its terms and definitions, can be found here: <https://addictovocab.org>.

direction. As I see it, this involves ensuring both its conformance with BFO and the principles of best practice, as well as the integrity of the substantive account of addiction that AddictO takes on through the definitions and relations it asserts. Third, a quick review of the AddictO ontology suggests that it is already an improvement on extant ontological representations of addiction and surrounding phenomena. The account of addiction and the ontology itself better capture the criteria from Section 5.2. Moreover, AddictO is the first ontology dedicated to bringing *addiction-relevant* entities and relations together in a systematic way. Consider the AddictO definition of ‘addiction’:

addiction =def. *A mental disposition towards repeated episodes of abnormally high levels of motivation to engage in a behaviour, acquired as a result of engaging in the behaviour, where the behaviour results in risk or occurrence of serious net harm.*

The most notable improvement is AddictO’s explicit classification of **ADDICTION** as *subtype_of* **DISPOSITION**, thereby avoiding the conflation of addiction with its behavioral manifestations (a common problem from above). AddictO also includes a distinct class called ‘addictive disorder’, which resembles Wakefield’s distinction between, on the one hand, addiction *qua* impairment of control over desires and motivation, and, on the other, the condition wherein addiction becomes a disease (which he also calls ‘addictive disorder’).⁴⁸⁴ As SC9 indicates, I think this is the right distinction to make.

However, for all its improvements over existing ontological work on addiction, AddictO is still not quite a complete and fully well-built ontology. It also falters on some of the above criteria. For instance, Figure 28 below shows that AddictO currently has 63 root terms that are simply pulled from other ontologies (likely for their relevance to addiction and surrounding phenomena, such as addiction treatment or research) and organized as a collection of sibling classes.

⁴⁸⁴ See Wakefield (2017a).

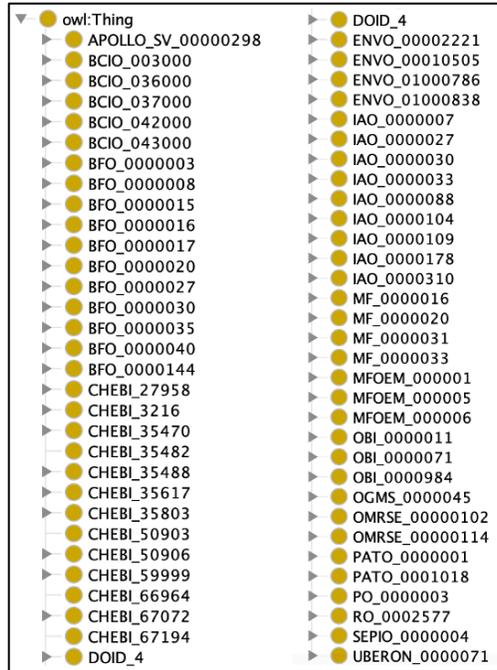


Figure 28: Top-Level (Root) Terms in AddictO

Furthermore, we can see from the definition above that AddictO violates the criteria regarding the harm condition and the historical condition. Another example of a flaw in AddictO is that the above-mentioned class **ADDICTIVE DISORDER** is currently a child of **HUMAN BEING**, a clear (though likely accidental) error in classification. Finally, consider that AddictO defines ‘behaviour attribute’ as “An element of a behaviour process *which can be used to characterise the behaviour*,” (emphasis added) which seems to violate realism because the entity here is defined in terms of our ability to know about it and use it in descriptions (such as theories). An element of a behavior, whatever that comes to, should not depend on our ability to discover, discuss, or theorize about it.

These criticisms are not meant to deride AddictO or its clear potential for improving on current work on the ontology of addiction. They are likely the result of the fact that this project has only just begun. For this and the reasons above, AddictO was not considered in the main evaluations above.

5.5 Where We Stand: Summarizing the Arguments of the Foregoing Chapters

We have come to the end of the project. Thus, I want to now bring things around full circle and summarize the defenses of the premises of my overarching argument. The argument is restated in pieces below, with the relevant summaries appearing after each premise.

Here is premise (1):

(1) *The addiction literature suffers from a serious problem of disunification.* [Chapter 1]

In Chapter 1, I pointed out a number of different kinds of problems facing the addiction literature. For instance, authors often provide incomplete, vague, ambiguous, or otherwise unclear definitions of their terms; authors sometimes provide no definition at all for ‘addiction’ or related terms they use like ‘disease’ or ‘compulsion’; authors often use the same term to mean different things or different terms to mean the same thing; and authors sometimes provide problematic definitions that make unfounded inferences, such as ‘disease’ being defined in terms of irrationality or entailing compulsion. Moreover, stalled debates and conceptual confusions abound. In general, the problem with the literature is that authors continually talk past one another, and moreover there is no established practice in the literature for remedying this problem.

These are all problems of disunification. Methodological unification allows researchers to effectively interact and communicate independently of the content of their respective views. A set of views are methodologically unified just in case, roughly, they have the appropriate means to successfully interact with and understand each other – a shared vocabulary, principles for controlling it, and so forth. Hence, given the evidence, the addiction literature is currently methodologically *dis*unified. Substantive unification, on the other hand, is about attempting to bring the content of different (often competing) views together under a single, unified framework. A set of views is substantively unified just in case there is some account the contents of which can be said to capture the others. We have seen the contents of numerous extant accounts throughout this and the preceding

chapters. Moreover, the overwhelming tendency to focus on a particular disciplinary approach (and corresponding aspect of addiction, such as the material basis in the brain, behavioral manifestations, and environmental influences) demonstrates why the blind men and the elephant metaphor has seemed so appropriate. As they stand, it is hard to see how any extant view could capture most, or even some, of the others on offer. Hence, given the evidence, the addiction literature is also currently *substantively* disunified.

Here is premise (2):

(2) *Ontology can help solve the addiction literature's problem of disunification.* [Chapter 2]

In Chapter 2, I laid out the main components of ontology, discussed some of the core principles of best practice for ontology building, and explained how applying this method to the addiction literature can help to solve its problems of disunification. We also saw that the need for ontology arises out of a context with two central features: scientific investigation of the world and heterogeneous data. The domain of addiction research clearly has both. Moreover, we saw two examples of the success of applying ontology to the biological and medical domains in order to facilitate unification: The Gene Ontology and the Ontology for General Medical Science. Finally, I argued *a la* Robert West, Janna Hastings, and colleagues that, with respect to unification and the application of ontology, there is no relevant difference between these domains (biology and medicine) and the domain of addiction research. Hence, ontology can help to solve the *addiction* literature's problems with disunification.

Here is premise (3):

(3) *The dispositionalist account of addiction is true and provides still further unification to the literature.* [Chapters 3 and 4]

In Chapter 3, I set the stage for defending this premise by laying out the dispositionalist account of addiction, involving desires, impaired control, and systematicity, and I provided the basic background modal metaphysic which this account assumes. This centered around the nature of dispositions, including triggering conditions, realizations, manifestations, material basis, and three senses of

reliability: strength, opportunity, and systematicity. I also argued that ontology is somewhat limited in its unifying power. Given certain theory-neutral and pluralistic principles like *perspectivalism*, *adequatism*, and *AAA*, ontology is capable of providing only methodological unification and not substantive unification. This left room for the dispositionalist account to pick up where ontology left off.

Thus, in Chapter 4, I argued that the dispositionalist account of addiction best explains key phenomena, such as the puzzle of addiction. Moreover, I argued that it can capture most extant competing accounts of addiction. This is because it is either consistent with these accounts (because, for instance, they are talking about the behavioral manifestations or the material basis of addiction) or the accounts are already getting at a disposition (though, typically implicitly and almost always only in a cursory manner without filling out the metaphysical details). I then provided reasons for thinking that the dispositionalist account's unifying power supports both its truth and its ability to substantively unify the literature. Finally, I offered a glimpse of how my account would be ontologized and explained how it is consistent with the methods of good ontology building (specifically, being BFO-conformant).

Here is premise (4):

- (4) *The dispositionalist account of addiction fares better than competitor accounts, both philosophical and ontological.* [Chapter 5]

The present chapter defended this comparative premise. I laid out and defended two sets of criteria for an account of addiction, one substantive (about the account itself) and one methodological (about using proper ontological principles). The former was used to evaluate a number of prominent views of addiction. Then both sets were used to evaluate representations of addiction in existing ontologies and other classification systems and vocabularies important to addiction research. All failed on at least some and often many of these criteria. Thus, the dispositionalist account not only helps to unify the literature, but it avoids many of the pitfalls that existing accounts fall into.

Here is premise (5):

- (5) *If (1)-(4), then addiction research ought to develop and implement an ontology of addiction and adopt a dispositionalist account of addiction. [trivial]*

This premise is straightforward and really quite trivial. Assume that the literature is methodologically disunified and that ontology can remedy this (though I have argued for both at length). It would seem to follow very easily that addiction research should incorporate the methods of ontology. Further, assume that addiction is, in fact, a disposition and that recognizing this can help to provide substantive unification to the literature (again, I have argued for both at length). It should likewise follow very easily that the dispositionalist account should be adopted. Thus, this bridge premise is on solid ground.

Hence, the arguments presented in the preceding chapters allows us to conclude the following:

- (6) *Hence, addiction research ought to develop and implement an ontology of addiction and adopt a dispositionalist account of addiction. [(1)-(5)]*

References

- Ainscough, T. S., McNeill, A., Strang, J., Calder, R., & Brose, L. S. (2017). Contingency management interventions for non-prescribed drug use during treatment for opiate addiction: A systematic review and meta-analysis. *Drug and Alcohol Dependence*, 178, 318–339.
- Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press.
- Ainslie, G. (2013). Intertemporal bargaining in addiction. *Frontiers in Psychiatry*, 4(63), 1–5.
- Ainslie, G. (2017). Palpating the elephant: Current theories of addiction in light of hyperbolic delay discounting. In N. Heather & G. Segal (Eds.), *Addiction and choice: Rethinking the relationship* (pp. 227–244). Oxford University Press.
- Ainslie, G. (2019). The piceoeconomics of addiction. In S. H. Ahmed & H. Pickard (Eds.), *The Routledge handbook of the philosophy and science of addiction* (pp. 34–44). Routledge.
- Ainslie, G., & Monterosso, J. (2003). Hyperbolic discounting as a factor in addiction. In R. E. Vuchinich & N. Heather (Eds.), *Choice, behavioural economics and addiction* (pp. 35–69). Elsevier.
- Alexander, B. (2010). *The globalisation of addiction: A study in poverty of the spirit*. Oxford University Press.
- Alexander, B., & Schweighofer, A. R. F. (1988). Defining “addiction.” *Canadian Psychology*, 29(2), 151–162.
- Allemang, D., & Hendler, J. A. (2011). *Semantic Web for the working ontologist: Effective modeling in RDFS and OWL* (2nd edition). Elsevier.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th edition). American Psychiatric Association.
- Andreou, C. (2007). Making a clean break: Addiction and Ulysses contracts. *Bioethics*, 22(1), 25–31.
- Aristotle. (2001). *The basic works of Aristotle* (R. McKeon, Ed.). Modern Library.
- Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with Basic Formal Ontology*. MIT Press.
- Arpaly, N., & Schroeder, T. (2014). *In praise of desire*. Oxford University Press.

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29.
- Austin, J. L. (1956). Ifs and cans. *Proceedings of the British Academy*, *42*(1), 109–132.
- Bada, M., Stevens, R., Goble, C., Gil, Y., Ashburner, M., Blake, J. A., Cherry, J. M., Harris, M., & Lewis, S. (2004). A short study on the success of the Gene Ontology. *Journal of Web Semantics*, *1*(2), 235–240.
- Ballantyne, J. C., & LaForge, S. K. (2007). Opioid dependence and addiction during opioid treatment of chronic pain. *Pain*, *129*(3), 235–255.
- Barcellos Almeida, M., & Farinelli, F. (2017). Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. *Journal of the Association for Information Science and Technology*, *68*(11), 2529–2542.
- Baumeister, R. (2018). *Self-regulation and self-control: Selected works of Roy Baumeister*. Routledge.
- Benn, P. (2007). Disease, addiction and the freedom to resist. *Philosophical Papers*, *36*(3), 465–481.
- Berczik, K., Szabó, A., Griffiths, M. D., Kurimay, T., Kun, B., Urbán, R., & Demetrovics, Z. (2012). Exercise addiction: Symptoms, diagnosis, epidemiology, and etiology. *Substance Use & Misuse*, *47*(4), 403–417.
- Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, *71*(8), 670–679.
- Bhikkhu, T. (2012). *Tittha Sutta: Sectarians (1) (Ud 6.4)*. Access to Insight (BCBS Edition). <http://www.accesstoinsight.org/tipitaka/kn/ud/ud.6.04.than.html>
- Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction (Abingdon, England)*, *96*(1), 73–86.

- Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology*, *146*(4), 447–454.
- Bird, M. (2015, October 12). There's a crazy conspiracy theory that the Rothschilds sank the Titanic to set up the Federal Reserve. *Business Insider*. <https://www.businessinsider.com/conspiracy-theory-that-the-rothschilds-and-federal-reserve-proponents-sank-the-titanic-2015-10>
- Bishop, M. A. (2015). *The good life: Unifying the philosophy and psychology of well-being*. Oxford University Press.
- Bittner, T., & Smith, B. (2003). A theory of granular partitions. In M. Worboys, M. Duckham, & M. Goodchild (Eds.), *Foundations of geographic information science* (pp. 117–149). CRC Press.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*(4), 542–573.
- Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, *39*(6), 683–724.
- Borghini, A., & Williams, N. E. (2008). A dispositional theory of possibility. *Dialectica*, *62*(1), 21–41.
- Bourget, D., & Chalmers, D. (2014). What do philosophers believe? *Philosophical Studies*, *170*(3), 465–500.
- Buckner, J. D., Heimberg, R. G., Ecker, A. H., & Vinci, C. (2013). A biopsychosocial model of social anxiety and substance use. *Depression and Anxiety*, *30*(3), 276–284.
- Buffalo Developers Group. (2020). *Natural language specification of BFO 2020*.
- Buttigieg, P., Morrison, N., Smith, B., Mungall, C. J., Lewis, S. E., & the ENVO Consortium. (2013). The Environment Ontology: Contextualising biological and biomedical entities. *Journal of Biomedical Semantics*, *4*(43), 1–9.
- Ceusters, W., & Smith, B. (2010). Foundations for a realist ontology of mental disease. *Journal of Biomedical Semantics*, *1*(10), 1–23.

- Ceusters, W., & Smith, B. (2015). Aboutness: Towards foundations for the Information Artifact Ontology. *Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO), CEUR 1515*, 1–5.
- Charland, L. C. (2012). The varieties of compulsion in addiction. *American Journal of Bioethics - Neuroscience*, 3(2), 50–51.
- Cline, B. (2016). Against deliberative indispensability as an independent guide to what there is. *Philosophical Studies*, 173(12), 3235–3254.
- Courtot, M. (2016, September 19). *Ontologies for life sciences: Examples from the Gene Ontology* [PowerPoint]. Earlham Institute Summer School on Bioinformatics, Norwich, UK. <https://www.slideshare.net/mcourtot/ontologies-for-life-sciences-examples-from-the-gene-ontology>
- Cowell, L. G., & Smith, B. (2010). Infectious Disease Ontology. In V. Sintchenko (Ed.), *Infectious disease informatics* (pp. 373–395). Springer.
- Cox, S., Hastings, J., West, R., & Notley, C. (2020). The case for development of an E-cigarette Ontology (E-CigO) to improve quality, efficiency and clarity in the conduct and interpretation of research. *Qeios*. <https://www.qeios.com/read/article/545>
- Dalley, J. W., Everitt, B. J., & Robbins, T. W. (2011). Impulsivity, compulsivity, and top-down cognitive control. *Neuron*, 69(4), 680–694.
- Dalrymple, T. (2008). *Romancing opiates: Pharmacological lies and the addiction bureaucracy*. Encounter Books.
- Davidson, D. (2001). *Essays on actions and events* (2nd ed). Oxford University Press.
- Davies, J. B. (1997). *The myth of addiction* (2nd edition). Harwood Academic Publishing.
- de Wit, H. (2009). Impulsivity as a determinant and consequence of drug use: A review of underlying processes. *Addiction Biology*, 14(1), 22–31.
- Dennett, D. (2004). *Freedom evolves*. Penguin Books.

- Dessimoz, C., & Škunca, N. (Eds.). (2017). *The Gene Ontology handbook* (Vol. 1446). Springer.
- Di Chiara, G. (1999). Drug addiction as dopamine-dependent associative learning disorder. *European Journal of Pharmacology*, 375(1–3), 13–30.
- Dill, B., & Holton, R. (2014). The addict in us all. *Frontiers in Psychiatry*, 5(139), 1–20.
- du Plessis, G. P. (2012). Toward an integral model of addiction: By means of integral methodological pluralism as a metatheoretical and integrative conceptual framework. *Journal of Integral Theory and Practice*, 7(3), 1–24.
- du Plessis, G. P. (2014b). An integral ontology of addiction: A multiple object as a continuum of ontological complexity. *Journal of Integral Theory and Practice*, 9(1), 38–54.
- du Plessis, G. P. (2018). *An integral foundation for addiction treatment: Beyond the biopsychosocial model*. Integral.
- Edelstein, E. L. (1973). A case of water dependence. *Addiction*, 68(4), 365–367.
- El-Sappagh, S., Franda, F., Ali, F., & Kwak, K.-S. (2018). SNOMED CT standard ontology based on the Ontology for General Medical Science. *BMC Medical Informatics and Decision Making*, 18(76), 1–19.
- Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, 8(11), 1481–1489.
- Everitt, B. J., & Robbins, T. W. (2016). Drug addiction: Updating actions to habits to compulsions ten years on. *Annual Review of Psychology*, 67, 23–50.
- Feit, N. (2017). Harm and the concept of medical disorder. *Theoretical Medicine and Bioethics*, 38(5), 367–385.
- Feit, N. (2020). Medical disorder, harm, and damage. *Theoretical Medicine and Bioethics*, 4(1), 39–52.
- Feldstein Ewing, S. W., Apodaca, T. R., & Gaume, J. (2016). Ambivalence: Prerequisite for success in motivational interviewing with adolescents? *Addiction*, 111(11), 1900–1907.
- Fischer, J. M. (1982). Responsibility and control. *The Journal of Philosophy*, 79(1), 24.

- Fischer, J. M., & Ravizza, M. (2000). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.
- Flanagan, O. (2013). The shame of addiction. *Frontiers in Psychiatry, 4*(120), 1–11.
- Flanagan, O. (2016). Willing addicts? Drinkers, dandies, druggies, and other Dionysians. In N. Heather & G. Segal (Eds.), *Addiction and choice: Rethinking the relationship* (pp. 66–81). Oxford University Press.
- FNIDA. (2017). *The Friends of the National Institute on Drug Abuse*. The Friends of NIDA. <http://www.thefriendsofnida.org/>
- Foddy, B. (2011). Addiction and its sciences-philosophy. *Addiction, 106*(1), 25–31.
- Foddy, B. (2017). Addiction: The pleasures and perils of operant behavior. In N. Heather & G. Segal, *Addiction and choice: Rethinking the relationship* (pp. 49–65). Oxford University Press.
- Foddy, B., & Savulescu, J. (2006). Addiction and autonomy: Can addicted people consent to the prescription of their drug of addiction? *Bioethics, 20*(1), 1–15.
- Foddy, B., & Savulescu, J. (2010a). A liberal account of addiction. *Philosophy, Psychiatry, & Psychology, 17*(1), 1–22.
- Foddy, B., & Savulescu, J. (2010b). Relating addiction to disease, disability, autonomy, and the good life. *Philosophy, Psychiatry, & Psychology, 17*(1), 35–42.
- Fragoso, G., de Coronado, S., Haber, M., Hartel, F., & Wright, L. (2004). Overview and utilization of the NCI Thesaurus. *Comparative and Functional Genomics, 5*(8), 648–654.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy, 68*(1), 5–20.
- Frankfurt, H. (1988). *The importance of what we care about: Philosophical essays*. Cambridge University Press.
- Franklin, C. (2015). Agent-causation, explanation, and akrasia: A reply to Levy's Hard Luck. *Criminal Law and Philosophy, 9*(4), 753–770.

- Froehlich, J. C., Fischer, S. M., Nicholson, E. R., Dilley, J. E., Filosa, N. J., Smith, T. N., & Rademacher, L. C. (2017). A combination of Naltrexone + Varenicline retards the expression of a genetic predisposition toward high alcohol drinking. *Alcoholism: Clinical and Experimental Research*, *41*(3), 644–652.
- The Gene Ontology Consortium. (2019). The Gene Ontology resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338.
- Ginet, C. (1990). *On action*. Cambridge University Press.
- Gkoutos, G. V., Schofield, P. N., & Hoehndorf, R. (2012). The Neurobehavior Ontology: An ontology for annotation and integration of behavior and behavioral phenotypes. In E. J. Chesler & M. A. Haendel (Eds.), *Bioinformatics of behavior: Part 1* (pp. 69–87). Elsevier.
- Goldberg, A. E. (2020). The (in)significance of the addiction debate. *Neuroethics*, *13*(3), 311–324.
- Goldstein, R. Z., & Volkow, N. (2002). Drug addiction and its underlying neurobiological basis: Neuroimaging evidence for the involvement of the frontal cortex. *The American Journal of Psychiatry*, *159*(10), 1642–1652.
- Gramlich, J. (2017, October 26). *Nearly half of Americans have a family member or close friend who's been addicted to drugs*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2017/10/26/nearly-half-of-americans-have-a-family-member-or-close-friend-whos-been-addicted-to-drugs/>
- Grodin, E. N., Steckler, L. E., & Momenan, R. (2016). Altered striatal response during effort-based valuation and motivation in alcohol-dependent individuals. *Alcohol and Alcoholism*, *51*(6), 638–646.
- Hall, W., & Weier, M. (2017). Lee Robins' studies of heroin use among US Vietnam veterans. *Addiction*, *112*(1), 176–180.
- Hanson, C. (2009). *Thinking about addiction: Hyperbolic discounting and responsible agency*. Rodopi.

- Hastings, J., Cox, S., West, R., & Notley, C. (2020). Addiction Ontology: Applying basic formal ontology in the addiction domain. *Qeios*. <https://www.qeios.com/read/HZHJIP>
- Hastings, J., Frishkoff, G. A., Smith, B., Jensen, M., Poldrack, R. A., Lomax, J., Bandrowski, A., Imam, F., Turner, J. A., & Martone, M. E. (2014). Interdisciplinary perspectives on the development, integration, and application of cognitive ontologies. *Frontiers in Neuroinformatics*, 8(62), 1–7.
- Hastings, J., le Novere, N., Ceusters, W., Mulligan, K., & Smith, B. (2012a). Wanting what we don't want to want: Representing addiction in interoperable bio-ontologies. *Proceedings of the Third International Conference on Biomedical Ontology*, 56–60.
- Hastings, J., & Schulz, S. (2012). Ontologies for human behavior analysis and their application to clinical data. *International Review of Neurobiology*, 103, 89–107.
- Hastings, J., Smith, B., Ceusters, W., Jensen, M., & Mulligan, K. (2012). Representing mental functioning: Ontologies for mental health and disease. In *ICBO 2012: 3rd International Conference on Biomedical Ontology* (pp. 1–5). University of Graz.
- He, Y., Yu, H., Ong, E., Wang, Y., Liu, Y., Huffman, A., Huang, H., Beverley, J., Hur, J., Yang, X., Chen, L., Omenn, G. S., Athey, B., & Smith, B. (2020). CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. *Scientific Data*, 7(1), 1–5.
- Heather, N. (2017a). Addiction as a form of akrasia. In N. Heather & G. Segal (Eds.), *Addiction and choice: Rethinking the relationship* (pp. 133–150). Oxford University Press.
- Heather, N. (2017b). On defining addiction. In N. Heather & G. Segal (Eds.), *Addiction and choice: Rethinking the relationship* (pp. 3–26). Oxford University Press.
- Heather, N. (2017c). Is the concept of compulsion useful in the explanation or description of addictive behaviour and experience? *Addictive Behaviors Reports*, 6, 15–38.

- Heather, N. (2020). The concept of akrasia as the foundation for a dual systems theory of addiction. *Behavioural Brain Research*, 390, 1–9.
- Heather, N., & Segal, G. (2015). Is addiction a myth? Donald Davidson’s solution to the problem of akrasia says not. *The International Journal of Alcohol and Drug Research*, 4(1), 77–83.
- Heil, J. (2012). *The universe as we find it*. Clarendon Press.
- Henden, E. (2013a). Addictive actions. *Philosophical Psychology*, 26(3), 362–382.
- Henden, E. (2013b). Heroin addiction and voluntary choice: The case of informed consent. *Bioethics*, 27(7), 395–401.
- Henden, E. (2016a). Addiction, compulsion, and weakness of the will: A dual-process perspective. In N. Heather & G. Segal (Eds.), *Addiction and choice: Rethinking the relationship* (pp. 116–132). Oxford University Press.
- Henden, E. (2016b). Addiction, voluntary choice, and informed consent: A reply to Uusitalo and Broers. *Bioethics*, 30(4), 293–298.
- Henden, E., Melberg, H. O., & Røgeberg, O. J. (2013). Addiction: Choice or compulsion? *Frontiers in Psychiatry*, 4(77), 1–11.
- Henderson, E. C. (2000). *Understanding addiction*. University Press of Mississippi.
- Herdova, M. (2015). Simply irresistible: Addiction, responsibility, and irresistible desires. *Journal of Cognition and Neuroethics*, 3(1), 195–216.
- Hershenov, D. (2020). A naturalist response to Kingma’s critique of naturalist accounts of disease. *Theoretical Medicine and Bioethics*, 41(2–3), 83–97.
- Heyman, G. (2009). *Addiction: A disorder of choice*. Harvard University Press.
- Heyman, G. (2013). Addiction and choice: Theory and new data. *Frontiers in Psychiatry*, 4(31), 1–5.

- Heyman, G. (2019). Deriving addiction: An analysis based on three elementary features of making choices. In S. H. Ahmed & H. Pickard (Eds.), *The Routledge handbook of the philosophy and science of addiction* (pp. 23–33). Routledge.
- Hodgson, J., Moore, K., Acri, T., & Treisman, G. J. (2020). *Integrative medicine for vulnerable populations: A clinical guide to working with chronic and comorbid medical disease, mental illness, and addiction*. Springer.
- Hoyt, C., Mungall, C., Vasilevsky, N., Domingo-Fernández, D., Healy, M., & Colluru, V. (2020). *Extension of roles in the ChEBI Ontology* [Preprint].
- Hutcheon, D., & Bevilacqua, M. (2010). Psychogenic polydipsia: A review of past and current interventions for treating psychiatric inpatients diagnosed with psychogenic polydipsia (PPD). *Annals of the American Psychotherapy Association, 13*(1), 1–28.
- Hyman, S. E. (2005). Addiction: A disease of learning and memory. *The American Journal of Psychiatry, 162*(8), 1414–1422.
- Hyman, S. E. (2007). The Neurobiology of addiction: Implications for voluntary control of behavior. *The American Journal of Bioethics, 7*(1), 8–11.
- Jacobs, J. D. (Ed.). (2017). *Causal powers*. Oxford University Press.
- Jonas, D. E., Amick, H. R., Feltner, C., Bobashev, G., Thomas, K., Wines, R., Kim, M. M., Shanahan, E., Gass, C. E., Rowe, C. J., & Garbutt, J. C. (2014). Pharmacotherapy for adults with alcohol use disorders in outpatient settings: A systematic review and meta-analysis. *Journal of the American Medical Association, 311*(18), 1889–1900.
- Kahneman, D. (2013). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292.
- Kalant, H. (1989). The nature of addiction: An analysis of the problem. In A. Goldstein (Ed.), *Molecular and cellular aspects of the drug addictions* (pp. 1–28). Springer.

- Kalant, H. (1994). Aaron's rod, or the blind men's elephant? *Addiction*, *89*, 13–14.
- Kalivas, P. W., & Volkow, N. (2005). The neural basis of addiction: A pathology of motivation and choice. *American Journal of Psychiatry*, *162*(8), 1403–1413.
- Kane, R. (1998). *The significance of free will*. Oxford University Press.
- Karila, L., Wéry, A., Weinstein, A., Cottencin, O., Petit, A., Reynaud, M., & Billieux, J. (2014). Sexual addiction or hypersexual disorder: Different terms for the same problem? A review of the literature. *Current Pharmaceutical Design*, *20*(25), 4012–4020.
- Kendler, K. S., & Myers, J. (2015). Addiction resistance: Definition, validation and association with mastery. *Drug and Alcohol Dependence*, *154*, 236–242.
- Kennett, J. (2013). Just say no? Addiction and the elements of self control. In N. Levy (Ed.), *Addiction and Self-Control* (pp. 144–164). Oxford University Press.
- Kershner, S., & Kelly, R. (Forthcoming). Responsibility-foundation: Still needed and still missing. *Science, Religion, and Culture*.
- Kincaid, H., & Sullivan, J. A. (2010). Medical models of addiction. In D. Ross, H. Kincaid, D. Spurrett, & P. Collins (Eds.), *What is addiction?* (pp. 352–376). MIT Press.
- Koob, G. F., Arends, M. A., & Le Moal, M. (2014). *Drugs, addiction, and the brain*. Elsevier.
- Koob, G. F., & Le Moal, M. (2006). *Neurobiology of addiction*. Elsevier.
- Koob, George. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, *24*(2), 97–129.
- Koob, George. (2013). Addiction is a reward deficit and stress surfeit disorder. *Frontiers in Psychiatry*, *4*(72), 1–18.
- Larsen, K. R., Michie, S., Hekler, E. B., Gibson, B., Spruijt-Metz, D., Ahern, D., Cole-Lewis, H., Ellis, R. J. B., Hesse, B., Moser, R. P., & Yi, J. (2017). Behavior change interventions: The potential of ontologies for advancing science and practice. *Journal of Behavioral Medicine*, *40*(1), 6–22.

- Leshner, A. (1999). Science-based views of drug addiction and its treatment. *Journal of the American Medical Association*, 282(14), 1314–1316.
- Leshner, A., & Koob, G. (1999). Drugs of abuse and the brain. *Proceedings of the Association of American Physicians*, 111(2), 99–108.
- Leshner, Alan I. (1997). Addiction is a brain disease, and it matters. *Science*, 278(5335), 45–47.
- Levy, N. (2006). Autonomy and addiction. *Canadian Journal of Philosophy*, 36(3), 427–447.
- Levy, N. (2011a). *Hard luck: How luck undermines free will and moral responsibility*. Oxford University Press.
- Levy, N. (2011b). Addiction, responsibility, and ego depletion. In J. Poland & G. Graham (Eds.), *Addiction and responsibility* (pp. 89–112). The MIT Press.
- Levy, N. (2013). Addiction is not a brain disease (and it matters). *Frontiers in Psychiatry*, 4(24), 1–7.
- Levy, N. (2014). Addiction as a disorder of belief. *Biology & Philosophy*, 29(3), 337–355.
- Levy, N. (2015). Addiction, autonomy, and informed consent: On and off the garden path. *Journal of Medicine and Philosophy*, jhv033.
- Levy, N. (2019). Addiction: The belief oscillation hypothesis. In H. Pickard & S. H. Ahmed (Eds.), *The Routledge handbook of philosophy and science of addiction* (pp. 54–63). Routledge.
- Lewis, D. K. (2001). *Counterfactuals*. Blackwell.
- Lewis, M. (2015). *The biology of desire: Why addiction is not a disease*. Public Affairs.
- Lewis, M. (2017). What evolution intended? Reply to Wakefield. *Neuroethics*, 10(1), 69–70.
- Lewis, S. E. (2017). The vision and challenges of the Gene Ontology. In C. Dessimoz & N. Škunca (Eds.), *The Gene Ontology handbook* (Vol. 1446, pp. 291–302). Springer.
- Ley, D. J. (2014). *Myth of sex addiction*. Rowman & Littlefield.
- Limbaugh, D. G. (2019). The harm of medical disorder as harm in the damage sense. *Theoretical Medicine and Bioethics*, 40(1), 1–19.

- Limbaugh, D. G., Jobst, L., Kasmier, D., Rudnicki, R., Llinas, J., & Smith, B. (2020). Ontology and cognitive outcomes. *Journal of Knowledge Structures and Systems*, 1(1), 3–22.
- Limbaugh, D. G., Kasmier, D., Ceusters, W., & Smith, B. (2019). Warranted Diagnosis. *Proceedings of the International Conference on Biomedical Ontology (ICBO)*, 1–10.
- Limbaugh, D. G., Kasmier, D., & Smith, B. (2019). *Primer on Capabilities* [Unpublished Manuscript].
- Linden, D. J. (2012). *The compass of pleasure: How our brains make fatty foods, orgasm, exercise, marijuana, generosity, vodka, learning, and gambling feel so good*. Penguin Books.
- Lowe, E. J. (2006). *The four-category ontology: A metaphysical foundation for natural science*. Oxford University Press.
- Lowe, E. J. (2010). *Personal agency: The metaphysics of mind and action*. Oxford University Press.
- Luske, H., Jackson, W., & Geronimi, C. (1955). *Lady and the Tramp*. Walt Disney Studios.
- Maisel, N. C., Blodgett, J. C., Wilbourne, P. L., Humphreys, K., & Finney, J. W. (2013). Meta-analysis of naltrexone and acamprosate for treating alcohol use disorders: When are these medications most helpful? *Addiction*, 108(2), 275–293.
- Malone, J., Holloway, E., Adamusiak, T., Kapushesky, M., Zheng, J., Kolesnikov, N., Zhukova, A., Brazma, A., & Parkinson, H. (2010). Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics*, 26(8), 1112–1118.
- Manley, D., & Wasserman, R. (2007). A gradable approach to dispositions. *Philosophical Quarterly*, 57(226), 68–75.
- Marlatt, G. A., Baer, J. S., Donovan, D. M., & Kivlahan, D. R. (1988). Addictive behaviors: Etiology and treatment. *Annual Review of Psychology*, 39(1), 223–252.
- Marmodoro, A. (2010). *The metaphysics of powers: Their grounding and their manifestations*. Routledge.
- Matentzoglou, N., Balhoff, J. P., Bello, S. M., Boerkoel, C. F., Bradford, Y. M., Carmody, L. C., Cooper, L. D., Grove, C. A., Harris, N. L., Köhler, S., Laporte, M.-A., Laudederkind, S. L. F., Lee, R.,

- Mazandu, G. K., McMurry, J. A., Mungall, C., Osumi-Sutherland, D., Pilgrim, C., Rageth, K., ... Haendel, M. A. (2018). *Phenotype Ontologies Traversing All The Organisms (POTATO) workshop aims to reconcile logical definitions across species.*
- McClure, J., Podos, J., & Richardson, H. N. (2014). Isolating the delay component of impulsive choice in adolescent rats. *Frontiers in Integrative Neuroscience*, 8(3), 1–9.
- McKittrick, J. (2018). *Dispositional pluralism*. Oxford University Press.
- Mele, A. (1992). *Irrationality: An essay on akrasia, self-deception, and self-control*. Oxford University Press.
- Mele, A. (2008). *Free will and luck*. Oxford University Press.
- Meule, A., & Gearhardt, A. N. (2014). Food addiction in the light of DSM-5. *Nutrients*, 6(9), 3653–3671.
- Michie, S., & Abraham, C. (2008). Advancing the science of behaviour change: A plea for scientific reporting. *Addiction*, 103(9), 1409–1410.
- Michie, S., & Johnston, M. (2017). Optimising the value of the evidence generated in implementation science: The use of ontologies to address the challenges. *Implementation Science*, 12(1), 131–134.
- Michie, S., Thomas, J., Johnston, M., Aonghusa, P. M., Shawe-Taylor, J., Kelly, M. P., Deleris, L. A., Finnerty, A. N., Marques, M. M., Norris, E., O'Mara-Eves, A., & West, R. (2017). The Human Behaviour-Change Project: Harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implementation Science*, 12(121), 1–12.
- Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change* (3rd edition). Guilford Press.
- Molnar, G., & Mumford, S. (2006). *Powers: A study in metaphysics*. Oxford University Press.
- Mumford, S. (2006). The ungrounded argument. *Synthese*, 149(3), 471–489.
- Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford University Press.

- Muth, M. (2001, July 19). Scientific conference focuses on ecstasy (MDMA): International experts meet to discuss the latest research and emerging trends. *NIDA News Release*.
<https://web.archive.org/web/20090827124917/http://drugabuse.gov/MedAdv/01/NR7-19.html>
- National Institutes of Health. (2005). *National Institute on Drug Abuse: A resource guide* (pp. 1–2).
<https://www.aamc.org/research/adhocgp/pdfs/nida.pdf>
- NIDA. (2020, July). *Glossary: Commonly used terms in addiction science*. National Institute on Drug Abuse.
<https://www.drugabuse.gov/publications/media-guide/glossary>
- Norris, E., Marques, M. M., Finnerty, A. N., Wright, A. J., West, R., Hastings, J., Williams, P., Carey, R. N., Kelly, M. P., Johnston, M., & Michie, S. (2020). Development of an Intervention Setting Ontology for behaviour change: Specifying where interventions take place. *Wellcome Open Research*, 5(124), 1–18.
- O'Brien, C. (2011). Addiction and dependence in DSM-V. *Addiction*, 106(5), 866–867.
- OMHA. (2017, June). *Glossary* [U.S. Department of Health and Human Services].
<https://www.hhs.gov/about/agencies/omha/filing-an-appeal/glossary/index.html>
- OMHA. (2020, August). *Opioid Crisis Statistics*. US Department of Health and Human Services: National Opioids Crisis. <https://www.hhs.gov/opioids/about-the-epidemic/opioid-crisis-statistics/index.html>
- Peele, S. (1987). Introduction to “Visions of Addiction”: The nature of the beast. *Journal of Drug Issues*, 17(1), 1–7.
- Pereboom, D. (cornell U. (2016). *Free will, agency, and meaning in life*. Oxford University Press.
- Pickard, H. (2015). Psychopathology and the ability to do otherwise. *Philosophy and Phenomenological Research*, 90(1), 135–163.
- Pickard, H. (2016). Denial in addiction. *Mind and Language*, 31(3), 277–299.

- Pickard, H. (2017). Responsibility without blame for addiction. *Neuroethics*, 10(1), 169–180.
- Pickard, H. (2019). The puzzle of addiction. In *The Routledge handbook of the philosophy and science of addiction* (pp. 9–22). Routledge.
- Pickard, H. (2020). What we're not talking about when we talk about addiction. *The Hasting's Center Report*, 50(4), 37–46.
- Pickard, H., & Ahmed, S. H. (2016). How do you know you have a drug problem? The role of knowledge of negative consequences in explaining drug choice in humans and rats. In N. Heather & G. Segal (Eds.), *Addiction and choice: Rethinking the relationship* (pp. 29–48). Oxford University Press.
- Pickard, H., & Ahmed, S. H. (Eds.). (2019). *The Routledge Handbook for the Philosophy and Science of Addiction*. Routledge.
- Pickard, H., & Pearce, S. (2013). Addiction in context. In N. Levy (Ed.), *Addiction and self-control: Perspectives from philosophy, psychology, and neuroscience* (pp. 165–189). Oxford University Press.
- Putnam, H. (1975). The meaning of “meaning.” *Language, Mind, and Knowledge*, 7(1), 131–193.
- Ralphs, J. A., de C Williams, A. C., Richardson, P. H., Pither, C. E., & Nicholas, M. K. (1994). Opiate reduction in chronic pain patients: A comparison of patient-controlled reduction and staff controlled cocktail methods: *Pain*, 56(3), 279–288.
- Redish, A. D. (2004). Addiction as a computational process gone awry. *Science*, 306(5703), 1944–1947.
- Reid, T. (1788). *Essays on the active powers of man* (K. Haakonssen & J. A. Harris, Eds.). Edinburgh University Press.
- Robbins, T., & Clark, L. (2015). Behavioral addictions. *Current Opinion in Neurobiology*, 30, 66–72.
- Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews*, 18(3), 247–291.

- Sanislow, C. A., Morris, S. E., Pacheco, J., & Cuthbert, B. N. (2020). The National Institute of Mental Health Research Domain Criteria: An alternative framework to guide psychopathology research. In J. R. Geddes, N. C. Andreasen, & G. M. Goodwin (Eds.), *New Oxford textbook of psychiatry* (3rd edition, pp. 62–72). Oxford University Press.
- Sarntivijai, S., Lin, Y., Xiang, Z., Meehan, T. F., Diehl, A. D., Vempati, U. D., Schürer, S. C., Pang, C., Malone, J., Parkinson, H., Liu, Y., Takatsuki, T., Saijo, K., Masuya, H., Nakamura, Y., Brush, M. H., Haendel, M. A., Zheng, J., Stoeckert, C. J., ... He, Y. (2014). CLO: The Cell Line Ontology. *Journal of Biomedical Semantics*, 5(37), 1–10.
- Sartorio, C. (2016). *Causation and free will*. Oxford University Press.
- Savage, S. R., Joranson, D. E., Covington, E. C., Schnoll, S. H., Heit, H. A., & Gilson, A. M. (2003). Definitions related to the medical use of opioids: Evolution towards universal agreement. *Journal of Pain and Symptom Management*, 26(1), 655–667.
- Schaler, J. A. (2000). *Addiction is a choice*. Open Court.
- Scheuermann, R. H., Ceusters, W., & Smith, B. (2009). Toward an ontological treatment of disease and diagnosis. *Summit on Translational Bioinformatics*, 116–120.
- Schriml, L. M., Mitraka, E., Munro, J., Tauber, B., Schor, M., Nickle, L., Felix, V., Jeng, L., Bearer, C., Lichenstein, R., Bisordi, K., Champion, N., Hyman, B., Kurland, D., Oates, C. P., Kibbey, S., Sreekumar, P., Le, C., Giglio, M., & Greene, C. (2019). Human Disease Ontology 2018 update: Classification, content and workflow expansion. *Nucleic Acids Research*, 47(D1), D955–D962.
- Schroeder, T. (2004). *Three faces of desire*. Oxford University Press.
- Schweighofer, N., Shishida, K., Han, C. E., Okamoto, Y., Tanaka, S. C., Yamawaki, S., & Doya, K. (2006). Humans Can Adopt Optimal Discounting Strategy under Real-Time Constraints. *PLoS Computational Biology*, 2(11), 1349–1356.

- Serecigni, J. G. (2015). Los antagonistas de los receptores opioides en el tratamiento del alcoholismo. *Adicciones*, 27(3), 214–230.
- Shaw, M., & Black, D. W. (2008). Internet addiction: Definition, assessment, epidemiology and clinical management. *CNS Drugs*, 22(5), 353–365.
- Sinnott-Armstrong, W., & Pickard, H. (2013). What is addiction? In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, & T. Thornton (Eds.), *The Oxford handbook of philosophy and psychiatry* (Vol. 1, pp. 851–864). Oxford University Press.
- Smith, B. (1995). On drawing lines on a map. In A. U. Frank & W. Kuhn (Eds.), *Spatial information theory: A theoretical basis for GIS* (Vol. 988, pp. 475–484). Springer.
- Smith, B. (2012). Classifying processes: An essay in applied ontology. *Ratio*, 25(4), 463–488.
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L., & Lewis, S. (2007). The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11), 1251–1255.
- Smith, B., & Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, 5(3–4), 139–188.
- Smith, B., & Ceusters, W. (2015). Biomarkers in the Ontology for General Medical Science. *Studies in Health Technology and Informatics*, 210, 155–159.
- Smith, C. L., & Eppig, J. T. (2009). The Mammalian Phenotype Ontology: Enabling robust annotation and comparative analysis. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(3), 390–399.
- Solomon, R. L. (1980). The opponent-process theory of acquired motivation: The costs of pleasure and the benefits of pain. *American Psychologist*, 35(8), 691–712.

- Sripada, C. (2017). Frankfurt's unwilling and willing addicts. *Mind*, 126(503), 781–815.
- Sripada, C., & Railton, P. (2018). Addiction and moral psychology. In H. Pickard & S. H. Ahmed (Eds.), *The Routledge handbook of philosophy and science of addiction* (pp. 63–76). Routledge.
- Stolberg, S., Kaplan, S., & Mervosh, S. (2020, June 3). C.D.C. test counting error leaves epidemiologists 'really baffled.' *New York Times*.
<https://www.nytimes.com/2020/05/22/us/politics/coronavirus-tests-cdc.html>
- Summers, J. S. (2015). What is wrong with addiction. *Philosophy, Psychiatry, & Psychology*, 22(1), 25–40.
- Sussman, S., & Sussman, A. N. (2011). Considering the definition of addiction. *International Journal of Environmental Research and Public Health*, 8(10), 4025–4038.
- The OBO Foundry. (2020a). *NCI Thesaurus OBO Edition*. The OBO Foundry.
<http://www.obofoundry.org/ontology/ncit.html>
- The OBO Foundry. (2020b). *Principles: Overview*. The OBO Foundry.
<http://obofoundry.org/principles/fp-000-summary.html>
- Trickey, E. (2018, August 4). Fact-checking QAnon conspiracy theories: Did J.P. Morgan sink the Titanic? *The Washington Post*.
<https://www.washingtonpost.com/news/retropolis/wp/2018/08/04/how-j-p-morgan-didnt-sink-the-titanic-and-other-qanon-conspiracy-theories-debunked/>
- Uusitalo, S. (2011). On addicts' moral responsibility and action. *Res Cogitans*, 8(1), 77–91.
- Uusitalo, S., & Broers, B. (2015). Rethinking informed consent in research on heroin-assisted treatment. *Bioethics*, 29(7), 462–469.
- Uusitalo, S., Salmela, M., & Nikkinen, J. (2013). Addiction, agency and affects: Philosophical perspectives. *Nordic Studies on Alcohol and Drugs*, 30(1–2), 33–50.
- Vetter, B. (2013). Multi-track dispositions. *Philosophical Quarterly*, 63(251), 330–352.
- Vetter, B. (2014). Dispositions without conditionals. *Mind*, 123(489), 129–156.

- Vetter, B. (2015). *Potentiality: From dispositions to modality*. Oxford University Press.
- Vetter, B. (2021). A plenitude of powers. *Synthese*, 198(S6), 1365–1385.
- Volkow, N. (2014). *Drugs, brains, and behavior: The science of addiction*. National Institute on Drug Abuse.
- Volkow, N., & Fowler, J. (2000). Addiction, a disease of compulsion and drive: Involvement of the orbitofrontal cortex. *Cerebral Cortex*, 10(3), 318–325.
- Volkow, N., & Koob, G. (2015). Brain disease model of addiction: Why is it so controversial? *The Lancet Psychiatry*, 2(8), 677–679.
- Volkow, N., Koob, G. F., & McLellan, A. T. (2016). Neurobiological advances from the brain disease model of addiction. *New England Journal of Medicine*, 374(4), 363–371.
- Volkow, N., & Morales, M. (2015). The brain on drugs: From reward to addiction. *Cell*, 162(4), 712–725.
- Wakefield, J. C. (1992). The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist*, 47(3), 16.
- Wakefield, J. C. (2014). The biostatistical theory versus the harmful dysfunction analysis, part 1: Is part-dysfunction a sufficient condition for medical disorder? *Journal of Medicine and Philosophy*, 39(6), 648–682.
- Wakefield, J. C. (2017a). Addiction and the concept of disorder, part 1: Why addiction is a medical disorder. *Neuroethics*, 10(1), 39–53.
- Wakefield, J. C. (2017b). Addiction and the concept of disorder, part 2: Is every mental disorder a brain disorder? *Neuroethics*, 10(1), 55–67.
- Wallace, R. J. (1999). Addiction as defect of the will: Some philosophical reflections. *Law and Philosophy*, 18(1), 621–654.
- Walters, G. D., & Gilbert, A. A. (2000). Defining addiction: Contrasting views of clients and experts. *Addiction Research*, 8(3), 211–220.

- Wang, L. (2017). Heterogeneous data and big data analytics. *Automatic Control and Information Sciences*, 3(1), 8–15.
- Watson, G. (1977). Skepticism about weakness of will. *Philosophical Review*, 86(3), 316–339.
- West, R., & Brown, J. (2013). *Theory of addiction* (2nd edition). Wiley Blackwell.
- West, R., Christmas, S., Hastings, J., & Michie, S. (2019a). Developing general models and theories of addiction. In S. H. Ahmed & H. Pickard (Eds.), *The Routledge handbook of the philosophy and science of addiction* (pp. 160–172). Routledge.
- West, R., Marsden, J., & Hastings, J. (2019a). Addiction Theories and Constructs: A new series. *Addiction*, 114(6), 955–956.
- West, R., Marsden, J., & Michie, S. (n.d.). *Improving of reporting addiction research* [PowerPoint].
- White, W. (2000). Addiction as a disease: Birth of a concept. *Counselor*, 1(1), 46–51.
- Williams, N. E. (2005). Static and dynamic dispositions. *Synthese*, 146(3), 303–324.
- Williams, N. E. (2009). The ungrounded argument is unfounded: A response to Mumford. *Synthese*, 170(1), 7–19.
- Williams, N. E. (2011). Dispositions and the argument from science. *Australasian Journal of Philosophy*, 89(1), 71–90.
- Williams, N. E. (2019). *The powers metaphysic* (First edition). Oxford University Press.
- Winston, S., & Seif, M. N. (2017). *Overcoming unwanted intrusive thoughts: A CBT-based guide to getting over frightening, obsessive, or disturbing thoughts*. New Harbinger Publications, Inc.
- Wise, R. A. (2000). Addiction becomes a brain disease. *Neuron*, 26(1), 27–33.
- Wolf, S. (1990). *Freedom within reason*. Oxford University Press.
- World Health Organization. (2020). *International classification of diseases and related health problems* (12th edition). World Health Organization.

- Wright, A. J., Norris, E., Finnerty, A. N., Marques, M. M., Johnston, M., Kelly, M. P., Hastings, J., West, R., & Michie, S. (2020). Ontologies relevant to behaviour change interventions: A method for their development. *Wellcome Open Research*, 5(126), 1–29.
- Xu, X., & Chaloupka, F. J. (2011). The effects of prices on alcohol use and its consequences. *Alcohol Research & Health*, 34(2), 236–245.
- Yau, Y. H. C., & Potenza, M. N. (2015). Gambling disorder and other behavioral addictions: Recognition and treatment. *Harvard Review of Psychiatry*, 23(2), 134–146.

Appendix A: Hyperbolic Discounting and Addiction

Everyone engages in temporal discounting – the further into the future some reward is, the more we discount the value of that reward, especially relative to a more immediate alternative reward. For instance, anyone would likely opt for \$100 now over \$20 now, but most people would also switch their preference and opt for \$20 now once the \$100 option was delayed far enough into the future (such as when the options are \$20 now or \$100 one year from now). The rate at which we discount delayed rewards relative to immediate ones is *hyperbolic*. This means that the change in expected value of a reward relative to the time delay is greater when the expected moment of reception is closer to the present. So, roughly, \$100 now is much, much better than \$100 tomorrow and \$100 next week, but \$100 next week is only somewhat better than \$100 in two weeks. See Figure 29 below for an illustration.

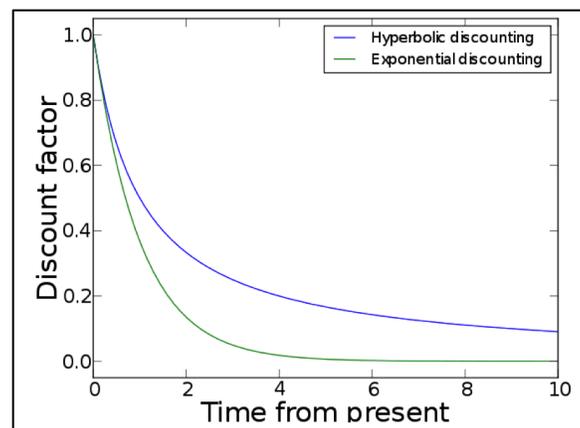


Figure 29: Illustration of a Hyperbolic Discounting Curve Relative to an Exponential Discounting Curve⁴⁸⁵

This helps to explain why we have a hard time staving off dessert (which tastes delicious now) despite our commitment to achieving long-term health or weight-loss goals (which is rewarding at some later time). More specifically, the hyperbolic nature of the discounting rate helps to explain why the dessert (relative to long-term weight-loss goals) seems so much more enticing at (and after!) dinner than it does when we contemplate these preferences absent any expectation of imminent reception.

⁴⁸⁵ Diagram taken from <https://neilpatel.com/blog/hyperbolic-discounting/>.

Indeed, this is why we can genuinely prefer the weight-loss goal (entailing no dessert) to the delicious taste of the dessert when a choice between them is not imminent, and then still undergo a preference shift once the choice for dessert is on the table, so-to-speak. See Figure 30 for an illustration.

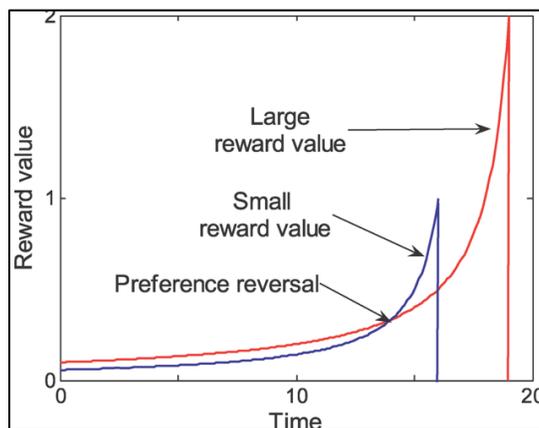


Figure 30: Illustration of a Preference Shift from Hyperbolic Discounting⁴⁸⁶

Though everyone does this, Ainslie helped to emphasize the fact that addicts engage in an *exaggerated* form of hyperbolic discounting.⁴⁸⁷ In short, addicts' discounting rates are also hyperbolic, but they discount delayed rewards at even faster rates. This is indicative of greater impulsivity, which is a commonly recognized feature of addiction. See Figure 31 for an illustration.

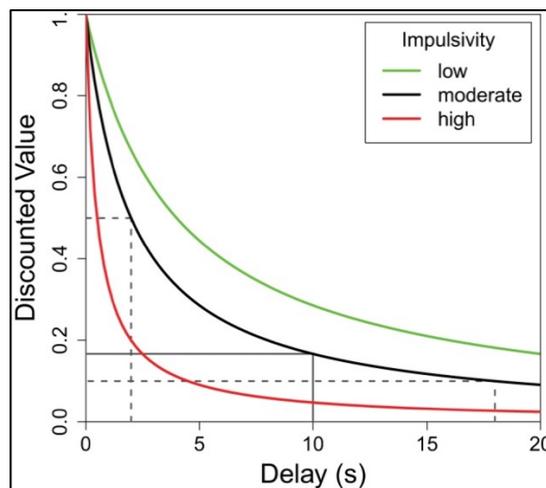


Figure 31: Difference in Discounting Rates Relative to Impulsivity⁴⁸⁸

⁴⁸⁶ Diagram taken from Schweighofer et al. (2006, p. 1351).

⁴⁸⁷ See Ainslie (2001).

⁴⁸⁸ Diagram taken from (McClure et al., 2014, p. 2).

Appendix B: OBO Foundry Principles of Best Practice for Ontology⁴⁸⁹

The OBO Foundry lists thirteen principles against which any ontology submitted to the Foundry is evaluated, and which the Foundry further recommends as good practice for any ontology building, whether it is intended to be a part of the Foundry or not. The principles and definitions are as follows:

1. **Open:** The ontology **MUST** be openly available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed in altered form under the original name or with the same identifiers.
2. **Common Format:** The ontology is made available in a common formal language in an accepted concrete syntax [such as OBO format, OWL, OWL 2, RDF/XML].
3. **URI/Identifier Space:** Each class and relation (property) in the ontology must have a unique URI identifier. The URI should be constructed from a base URI, a prefix that is unique within the Foundry (e.g. GO, CHEBI, CL) and a local identifier (e.g. 0000001). The local identifier should not consist of labels or mnemonics meaningful to humans.
4. **Versioning:** The ontology provider has documented procedures for versioning the ontology, and different versions of ontology are marked, stored, and officially released [such as on GitHub].
5. **Scope:** The scope of an ontology is the extent of the domain or subject matter it intends to cover. The ontology must have a clearly specified scope and content that adheres to that scope.
6. **Textual Definitions:** The ontology has textual definitions for the majority of its classes and for top level terms in particular [that is, they are unique and in Aristotelian form: **A** *is_a* **B** *that C*].
7. **Relations:** Relations should be reused [when possible] from the Relations Ontology (RO).
8. **Documentation:** The owners of the ontology should strive to provide as much documentation as possible. The documentation should detail the different processes specific to an ontology life cycle and target various audiences (users or developers).
9. **Documentation of Plurality of Users:** The ontology developers should document that the ontology is used by multiple independent people or organizations.

⁴⁸⁹ The following 13 principles are taken *verbatim* from <http://obofoundry.org/principles/fp-000-summary.html>. However, bracketed text has been added for further clarification.

10. **Commitment to Collaboration:** OBO Foundry ontology development, in common with many other standards-oriented scientific activities, should be carried out in a collaborative fashion.
11. **Locus of Authority:** There should be a person who is responsible for communications between the community and the ontology developers, for communicating with the Foundry on all Foundry-related matters, for mediating discussions involving maintenance in the light of scientific advance, and for ensuring that all user feedback is addressed.
12. **Naming Conventions:** Adherence to common naming conventions is more than just a matter of aesthetics. Such conventions provide guidance to ontology creators, help developers avoid flaws and inaccuracies when editing, and especially when interlinking, ontologies. Common naming conventions will also assist consumers of ontologies to more readily understand what meanings were intended by the authors of ontologies used in annotating bodies of data [such as unambiguous, plain English, unique, one label for each entity, and so on].⁴⁹⁰
13. **Maintenance:** The ontology needs to reflect changes in scientific consensus to remain accurate over time [such as monitor literature for changes, dissent, consensus, and so on].

⁴⁹⁰ This definition is taken from Schober et al. (2009).

Appendix C: BFO Principles of Best Practice for Ontology

Barry Smith and colleagues identify eight principles for building ontologies with BFO for specific domains, which have some overlap with the OBO Foundry principles (Appendix B). They are as follows:⁴⁹¹

1. **Realism:** the goal of an ontology is to describe reality.
2. **Perspectivalism:** there are multiple accurate descriptions of reality.
3. **Fallibilism:** ontologies, like scientific theories, are revisable in light of new discoveries.
4. **Adequatism:** the entities in a given domain should be taken seriously on their own terms, not viewed as reducible to other kinds of entities.
5. **The Principle of Reuse:** existing ontologies should be treated as benchmarks and reused whenever possible in building ontologies for new domains.
6. **The Ontology Design Process Should Balance Utility and Realism:** sacrificing realism to address considerations of short-term utility when building an ontology may be at the detriment of the ontology's longer-term utility.
7. **The Ontology Design Process Is Open-Ended:** scientific ontologies will always be subject to the need for update in light of advances in knowledge; ontology design, maintenance, and updating is an ongoing process.
8. **The Principle of Low-Hanging Fruit:** in ontology design, begin with the features of the relevant domain that are easiest to understand and define, then work outward to more complex and controversial features.

In addition, Smith and colleagues identify another twenty-five principles for ontology design, which are more specifically aimed at terminology, definitions, and classification. These, too, have some overlap with the OBO Foundry, and they are as follows:⁴⁹²

Terminology:

1. **Scientific Terms:** Include in the terminology terms used by influential groups of scientists for the most important types of entities in the domain to be represented.

⁴⁹¹ The following 8 principles are from Arp, Smith, & Spear (2015, p. 50).

⁴⁹² The following 25 principles are summarized from Arp, Smith, & Spear (2015, Ch. 4). The definitions are directly from the text, but the labels and bracketed text have been added for further clarification.

2. **Consistency with Subject Matter Experts:** Strive to ensure maximal consensus with the terminological usage of scientists in the relevant discipline. This may well involve working with domain experts, for instance in negotiating terminological compromises.
3. **Synonyms Where Entities Overlap but Terminology Doesn't:** Identify areas of disciplinary overlap where terminological usage is not consistent. Look for and keep track of synonyms for terms already in the terminology list from these areas.
4. **Use Extant Resources If Possible:** Don't reinvent the wheel. In term selection, stay as close as possible to the usage of actual domain experts. In terminology construction and ontology design, make use of as many existing resources (terminologies and ontologies) as possible.
5. **Singular Nouns:** Use singular nouns [such as using 'dog' instead of 'dogs'].
6. **Lowercase Italics for Nouns:** Use lowercase italic format for common nouns.
7. **No Acronyms:** Avoid acronyms.
8. **Use Unique URI:** Associate each term in the ontology with a unique alphanumeric identifier.
9. **Term Univocity:** Ensure univocity of terms [for instance, do not use 'bank' twice if possible].
10. **Relational Univocity:** Ensure univocity of relational expressions.
11. **No Mass Nouns:** Avoid mass nouns [such as using 'portion of sand' instead of 'sand'].
12. **General is not Particular:** Distinguish the general from the particular.

Definitions:

13. **Definitions Everywhere but Root Node:** Provide all non-root terms with definitions.
14. **Aristotelian Definitions:** Use Aristotelian definitions [**A** *is_a* **B** *that Cs*].
15. **Essentialist Definitions:** Use essential features in defining terms [**B** and *that Cs* are essential to being **A**, where being a **B** is shared by all of **A**'s siblings and *that Cs* differentiates **A** from all of its siblings].
16. **Top-Down:** Start with the most general terms in your domain.
17. **Non-Circular:** Avoid circularity in defining terms [for instance, do not use the defined term or one of its cognates in the definition].
18. **Simplicity in Definitions:** To ensure the intelligibility of definitions, use simpler terms [in the definition] than the term you are defining.
19. **Sparse Ontology:** Do not create terms for universals through logical combination [for instance, avoid classes like **DISEASE OR DISORDER** and **DOGS AND CATS**].

20. **Meaningful Substitutability of Terms with Definitions:** Definitions should be unpackable [that is, you can substitute the definition for the term in any sentence and retain the meaning].

Classification:

21. **Taxonomic Backbone:** Structure every ontology around a backbone *is_a* hierarchy.
22. **Complete:** Ensure *is_a* completeness [that is, every non-root entity has a parent].
23. **Single Inheritance:** Ensure asserted single inheritance [that is, have only one *is_a* relation per entity].
24. **Open World Assumption:** Both developers and users of an ontology should respect the open-world assumption...which means (here) that each ontology is built in a flexible manner to allow extension and correction, and is never put forward as providing a complete assay of the domain in question.
25. **Realism:** Adhere to the rule of objectivity, which means: describe what exists in reality, not what is known about what exists in reality.