

LOGICAL FORM AND THE LIMITS OF THOUGHT

by

Manish Oza

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto

© Copyright 2020 by Manish Oza

Abstract

Logical Form and the Limits of Thought

Manish Oza

Doctor of Philosophy

Graduate Department of Philosophy

University of Toronto

2020

What is the relation of logic to thinking? My dissertation offers a new argument for the claim that logic is constitutive of thinking in the following sense: representational activity counts as thinking only if it manifests sensitivity to logical rules. In short, thinking has to be minimally logical.

An account of thinking has to allow for our freedom to question or revise our commitments – even seemingly obvious conceptual connections – without loss of understanding. This freedom, I argue, requires that thinkers have general abilities to respond to support and tension among their thoughts. And these abilities are constituted by following logical rules. So thinkers have to follow logical rules.

But there isn't just one correct logic for thinking. I show that my view is consistent with logical pluralism: there are a range of correct logics, any one of which a thinker might follow. A logic for thinking does, however, have to contain certain minimal principles: Modus Ponens and Non-Contradiction, and perhaps others.

We follow logical rules by exercising logical capacities, which display a distinctive first-person/third-person asymmetry: a subject can find the instances of a rule compelling without seeing them as instances of a rule. As a result, there are two limits on illogical thinking. First, thinkers have to tend to find instances of logical rules compelling. Second, thinkers can't think in obviously illogical ways. So thinking has to be logical – but not perfectly so.

When we try to think, but fail, we produce nonsense. But our failures to think are often subjectively indistinguishable from thinking. To explain how this occurs, I offer an account of nonsense. To be under the illusion that some nonsense makes sense is to enter a pretence that the nonsense is meaningful. Our use of nonsense within the pretence relies on the role of logical form in understanding.

Finally, while the normativity of logic doesn't fall directly out of logical constitutivism, it's possible to build an attractive account of logical normativity which has logical constitutivism as an integral part. I argue that thinking is necessary for human flourishing, and that this is the source of logical normativity.

Acknowledgements

I have a lot of people to thank. As supervisor, Gurpreet Rattan was a model of open-mindedness and a font of new ideas. I couldn't have written this without his support, unstinting enthusiasm and wise advice. Thanks also to Philip Kremer, for teaching me logic; to Nick Stang, for guiding me through Kant and Hegel; and to both, for years of tough questions. In 2018, I had the chance to work with Florian Steinberger and Corine Besson in London; their hospitality showed me what it means to belong to an intellectual community. Special thanks go to Edward Kanterian for a few words in the margin of a tutorial essay on Husserl, in Oxford around 2009. I didn't know it at the time, but that was the seed of a dissertation!

I wouldn't have survived graduate school without Lisa Doerksen and John Bunke. Weekly conversations with them in the Green Beanery shaped how I think about everything. I'm grateful to John for reading the Introduction and chapter 4 on short notice, and to Moritz Bodner for reading chapter 2. I'd also particularly like to thank Julia Jael Smith and Jessica Wright, who've read and commented on nearly every chapter at some point in the last few years.

I've learned a huge amount from friends and colleagues in Toronto and elsewhere, including, in no particular order, Emily Perry, Josh Brandt, Prach Panchakunathorn, James Gillard, Lochin Brouillard, Catherine Rioux, Atoosa Kasirzadeh, David Dyzenhaus, Arthur Ripstein, Imogen Dickie, Dominic and Tashi Alford-Duguid, G. Anthony Bruno, Rachel O'Keefe, Adam Dobkin, Nir Av-Gay, Bernie Katz, Danny Goldstick, Mark Fortney, Jessica Leech, Carlo Nicolai, Clinton Tolley, AW Moore, Damian Melamedoff, Mason Westfall, Melissa Rees, Amogh Sahu, Sophia Arbeiter, Nader Shoaibi, Zuzanna Jusińska, Antonina Jamrozik, Ngozi Okidegbe, Anne-Sophie Ouellet, Steve Coyne, Dave Suarez, Mike Blezy, Aaron Henry, Zain Raza, Hamish Russell, Rory Harder and Seyed Yarandi.

I'd like to thank all the students I've taught at U of T for their patience and curiosity, and for teaching me how to explain things. I hope it shows. I also benefited from discussions of my work at the Ontario Quebec Hegel Organization, the American Philosophy Association, the Canadian Philosophy Association, the Wiener Forum für Analytische Philosophie, the Society for the Study of the History of Analytic Philosophy, the LEM Seminar at the Institute of Philosophy in London and the U of T Grad Forum. The work was supported by funding from SSHRC and OGS. Of course, all remaining errors are my own.

Hegel wrote that to study logic is 'to dwell and to labor in [a] realm of shadows'. Luckily, I had people to remind me to get out in the sun once in a while: my friends, my parents Amit and Sangeeta, my sister Janika, all my relatives, Maria Papachristos and her family and my dearest Richa. I can't express how grateful I am. This thesis is dedicated to the memory of Thomas Papachristos.

Contents

0	Introduction	1
0.1	Overview	1
0.2	Normativism and illogicality	2
0.2.1	The problem of illogical thinking	3
0.2.2	Back to Kant, again	5
0.3	Logic as constitutive of thinking	6
0.3.1	What I mean by ‘logic’	9
0.4	Dissertation outline	12
I	Developing logical constitutivism	15
1	Why thinking needs logic	16
1.1	Introduction	16
1.1.1	Terminology	16
1.1.2	Thesis and argument	18
1.2	The holistic constraint on understanding	19
1.2.1	Understanding requires a setting	20
1.2.2	Deviant understanding	25
1.3	General rational abilities	28
1.3.1	Two unsuccessful explanations of holism	29
1.3.2	A better explanation	31
1.4	The expressive role of logic	33
1.4.1	The conditional and negation	34
1.4.2	Response to Brandom	39
1.5	Quinean reprise	41
1.6	Conclusion	46
2	Disagreement and pluralism	47
2.1	Introduction	47
2.2	Monism and Disagreement	51
2.2.1	Type A disagreement: conflicting beliefs about logic	52
2.2.2	Type B disagreement: using conflicting logics	56
2.3	Pluralism and Disagreement	58
2.3.1	Domain variance	58
2.3.2	The minimal kit	62
2.3.3	Recap	66
2.4	Which logics contain MP and LNC?	67

3	Logical capacities	78
3.1	Introduction	78
3.2	Cognitivism	79
3.3	The causal-explanatory role of logic	83
3.3.1	Dogramaci’s associationism and the sensitivity condition . . .	83
3.3.2	The simple dispositional account	85
3.4	The first-personal character of inference	86
3.5	Logical capacities and tacit knowledge	88
3.5.1	Do thinkers have logical knowledge?	93
3.5.2	Logical capacities and separability	94
4	On the limits of illogical thinking	99
4.1	Logic as constitutive of thinking	99
4.2	The first limit on illogicality	100
4.2.1	The vagueness of <i>thinking</i>	103
4.2.2	The metaphysical reality of thinking	105
4.3	The second limit on illogicality	107
4.3.1	What makes a contradiction obvious?	109
4.3.2	<i>Reductio</i> proofs	111
4.4	Can we be certain someone isn’t thinking?	111
II	Consequences of constitutivism	113
5	Nonsense: a user’s guide	114
5.1	Introduction	114
5.2	The engagement constraint	116
5.2.1	Reasoning with nonsense	116
5.2.2	Nonsense-attributions	118
5.2.3	The minimalist account	120
5.3	The austerity constraint	124
5.3.1	Nonsense thoughts	125
5.3.2	What’s wrong with nonsense?	125
5.4	A pretence account of nonsense	128
5.4.1	Pretence and make-believe	128
5.4.2	Logical form and partial understanding	130
5.4.3	Pretentious nonsense: the general idea	132
5.4.4	Examples of pretentious nonsense	133
5.4.5	Objections and replies	136
5.5	Conclusion. Is there a transparent level of sense?	138

6	The value of thinking and the normativity of logic	140
6.1	Introduction	140
6.2	Logical self-constitution	144
6.2.1	The constitutivist account of practical reason	145
6.2.2	The constitutivist account of logical normativity	146
6.3	A problem for constitutivism and two unsuccessful responses	148
6.3.1	The absence of value problem	148
6.3.2	The dialectical response	149
6.3.3	The need to think	150
6.4	The value of thinking	152
6.4.1	Possible views of the value of thinking	153
6.4.2	Thinking and human flourishing	155
6.5	Vexing issues about logical normativity	159
6.5.1	Three questions about logical rules	160
6.5.2	The implausible consequences of the Simple Formulation	161
6.5.3	Bridge principles	164
	Bibliography	170

0 Introduction

... and for all her illegalities
an enemy of misrule.

Anne Carson

0.1 Overview

What is the relation of logic to thinking? Why should we think logically, and what happens if we don't? Is it even possible to think illogically?

These are the questions this dissertation seeks to answer. They aren't new questions, but my answer is different from the ones that contemporary philosophers tend to accept. In short, my answer is that following logical rules makes it possible for us to think, with all the conceptual flexibility that thinking involves. One aspect of this is that logic is the *form* of thinking: its skeleton, the internal structure that holds it up. Another aspect is that logic sets *limits* to thinking, drawing a line between thinking and non-thinking.

An analogy may help here. Philosophers sometimes talk about logic as 'the laws of thought', and it's useful to reflect on which laws they have in mind. Often, they seem to be thinking of the criminal law, which seeks to impose restrictions on preexisting activities, like stealing or killing. While the analogy isn't perfect, I think logic is better compared to the law of contract. Contract law makes it possible for us to do new things – to make various sorts of binding promises. It is power-conferring or facilitative.¹ Like the law of contract, logic sets conditions that we have to meet in order to do something; the 'penalty' for falling too far from these conditions is that we failed to do the thing the rules make possible.

Of course, this is all very imprecise. Making it precise will take up the next four chapters. My limited aim in this Introduction is to characterize my view in an

¹The contrast is from Hart (1961, 27): 'The criminal law is something which we either obey or disobey and what its rules require is spoken of as a 'duty'. ... Legal rules defining the ways in which valid contracts or wills or marriages are made do not require persons to act in certain ways whether they wish to or not. Such laws do not impose duties or obligations. Instead, they provide individuals with *facilities* for realizing their wishes...'

intuitive way and locate it in a historical context. In 0.2, I'll set out the dominant view of the relation of logic to thinking, make an objection and comment on the deeper source of the problem. In 0.3, I'll introduce my own view, highlighting some ways that it differs from other views and clarifying what I mean by 'logic'. In 0.4, I'll outline the rest of the dissertation.

0.2 Normativism and illogicality

In this section, I'll set out the dominant view of the relation of logic to thinking. I call this view *normativism*, following Nunez (2018).

Normativism Representational activity counts as thinking only if it is evaluable by logical rules.

To be clear, I have a lot of common ground with normativists. They hold that thinking is a distinct kind of representational activity – distinct from perceiving, sensing, or the mere information-processing of a thermostat. Thinking is the activity composed of acts of judging and inferring, acts whose contents are propositions or *thoughts*. They argue that at least part of what makes an activity thinking is its relation to logic. More precisely, they argue that there's some relation a representational activity must bear to logic in order to count as thinking. I agree with both of these claims; my disagreement with normativists is about *what* the relation is between logic and thinking.

Normativism originates in John MacFarlane's (2000; 2002) interpretation of Kant's philosophy of logic. Here is MacFarlane spelling out the view:

What makes [something] a thought is not that it *conforms to* the laws of logic, but that the laws of logic are normative for it. To say that the laws of logic are norms for thought as such, then, is not to say that it is impossible to think illogically, but only that it is impossible to think illogically and be thinking *correctly*. (2000, 54)

And here is Jessica Leech:

There are some normative laws, evaluability in light of which is constitutive of thought. That's just what thought *is*: a mental activity which is subject to rules of a peculiar kind. (2015, 17)

Normativists say that the relation of logic to thinking is that thinking is evaluable by logical standards; an activity that isn't up for logical assessment isn't thinking

at all. They emphasize, however, that thinking doesn't have to *conform* to those standards. Evaluability exhausts the relation of logic to thinking.

There are different ways of making normativism precise. One set of options is about which role logic plays. Steinberger (2019c) distinguishes three roles here: logic might provide 'directives' (standards that ought to guide subjects in their thinking), 'evaluations' (standards such that a subject's thinking goes well or poorly depending on the subject's conformity to the standards) or 'appraisals' (standards that we use to praise or criticize other subjects' thinking). Different norms might be appropriate for different roles.

The second set of options is about exactly what logic requires of thinkers. These options can be mapped out as a space of different 'bridge principles' (MacFarlane 2004; Field 2009; Steinberger 2019a). A bridge principle is a statement of the form

Bridge Principle If Γ entails q , [normative statement about attitudes to Γ and q].

As MacFarlane shows, we can generate different bridge principles by (among other things) varying the type of deontic operator, the scope of the operator, whether the obligation applies to accepting the consequence or not accepting its negation, or whether the obligation is conditional on knowing that the entailment holds. Here are some examples:

- If Γ entails q , then if you judge Γ , you ought to judge q .
- If Γ entails q , then if you judge Γ , you have reason not to judge *not- q* .
- If Γ entails q , then you ought to ensure that if you judge Γ , you judge q .
- If you know that Γ entails q , then if you judge Γ , you may judge q .

However, I'm not going to spend any more time on these different options, because any view on which such a principle exhausts the relation of logic to thinking faces a problem.

0.2.1 The problem of illogical thinking

The problem is that, if normativism is right, then thinking can be illogical to an arbitrary degree.

On the normativist view, the fact that some act of thinking violated a logical rule typically means that it was bad or prohibited. But no number of claims that certain acts of thinking were bad or prohibited add up to a claim that they were not acts of thinking at all. So no matter how often or how badly a subject violates

logical rules – even if every content the subject judges is logically false – the subject is still thinking.

You might think that I've left out part of the view. For example, Leech doesn't only say that thought counts as right or wrong in light of logical standards, but that it must count as right or wrong '[i]n order to count as thinking *at all*' (2015, 15). So an act will only count as thinking if it is evaluable by logical standards. Logical laws are, in Leech's terms, 'constitutive-normative' laws, which 'separate the Fs from the non-Fs, not in terms of whether or not something conforms to the law, but in terms of whether something is *subject to* or *evaluable in light of* the law' (2015, 2). Doesn't this rule out the worst cases of illogical thinking?

To assess this response, we have to consider what makes an act evaluable by logical standards. Not everything is, after all. It makes no sense, for example, to ask whether my cup of coffee is logical or illogical: the cup of coffee is simply not apt for evaluation by the standards of logic. So some things are apt for logical evaluation and others are not. And it can't be a brute fact that certain things and not others are apt for logical evaluation. There must be some explanation of why they are apt.

Why shouldn't a normativist say it's a brute fact which things are apt for logical evaluation? This is an unattractive claim – for one thing, it seems to leave potential explanations on the table. Judgments and sentences contain concepts or terms that express concepts; the coffee cup does not. Judgments and sentences have logical structure that determines their truth-conditions; the coffee cup does not. These facts suggest potential explanations for why judgments and sentences are apt for logical evaluation and the coffee cup is not. So to propose that the difference in evaluability is a brute fact would be to give up the investigation too early.

Let's say that an act is apt for logical evaluation if and only if it meets some condition C. Given that, for the normativist, an act must be apt for logical evaluation in order to be thinking, it follows that an act that fails to meet C is not thinking. For present purposes, we don't need to know what C is: whatever it is, it faces a dilemma. Either C rules out arbitrarily illogical thinking or it doesn't, and neither option is acceptable.

If C rules out arbitrarily illogical thinking, normativism doesn't exhaust the relation of logic to thinking. Normativism might still be true, in that case, but it would leave out the most fundamental point: an act has to *be* logical, not just be *evaluable* by logic, to be an act of thinking. So if C rules out arbitrarily illogical thinking, normativism doesn't capture the most basic way that logic relates to

thinking as it purports to do.

If C doesn't rule out illogical thinking – if C can be met by activity which violates logical rules to whatever degree – then thinking can be illogical to whatever degree.² Since the normativist purports to capture the most basic way that logic relates to thinking, this is the way they should go: they should accept the possibility of arbitrarily illogical thinking.

You might say: what's the problem? We make mistakes sometimes, so why couldn't we always make them?

When I imagine someone whose every judgment has the form of a contradiction – someone who looks out the window and judges that it is and isn't raining, that they should and shouldn't bring an umbrella, that the rain will and won't end in an hour – I'm not inclined to describe them as thinking. First, their 'judgments' have no meaningful connection to their evidence. Whether their evidence supports p or $not-p$, they will judge p and $not-p$. Second, their 'judgments' have no real expression in their conduct: we're never able to say that their actions make sense given what they believe. Third, their 'judgments' are never related to each other. If each of their judgments is contradictory, then none of their judgments is rationally supported by any of the others (although they might entail each other by the rule of Explosion). But without such links to evidence, conduct and other judgments, it's not clear how their judgments can have any content.

Still, I don't have a decisive argument against this possibility. All I will say is that the notion of such a thinker is counterintuitive, and that we shouldn't accept it unless forced to.

0.2.2 Back to Kant, again

Normativists trace their view back to Kant's notion of form. Kant said that logical rules are the 'form of thinking' (*Critique of Pure Reason* A54/B78) and 'the form of the understanding', which is the faculty of thinking (A56/B80). MacFarlane argues that the core of what Kant meant by this is that logical rules are *constitutive norms* for thinking, in the sense that an activity is thinking only if it is evaluable by logical rules.

²Of course, you might hold that illogical activity meets C, but violates some other necessary condition for thinking. This position would supplement normativism with a requirement that thinking conform to logic to some minimal degree. I don't think this position is *wrong*, but in substance, it's not a supplementation of normativism but a version of my own view.

Above, I argued that this position is philosophically unsatisfactory. But it's also unsatisfactory as a reading of Kant.³ Despite the fact that MacFarlane set out to investigate what it means to say logic is formal, it remains unclear what the idea of constitutive norms has to do with form or formality. The form of a thing is what unifies its elements and makes it the thing it is (Korsgaard 1996, 107): say, the way pieces of wood are put together to make a bookshelf. And what makes something a bookshelf isn't being evaluable by the standards for bookshelves, but conforming to them to some degree; something which utterly fails to conform to the form of a bookshelf isn't a bookshelf at all.

These two defects in normativism are connected. If logic is the structure of thinking, then thinking has to *be* logical, at least to some degree, not just be *evaluable* by logic. The reason why normativism leaves room for arbitrarily illogical thinking is that it doesn't take logic to be the structure of thinking. This connection is drawn by Hilary Putnam in a suggestive passage:

Logic is not a description of what holds true in “metaphysically possible worlds”, to use Kripke’s phrase. It is a doctrine of the *form of coherent thought*. Even if I think of what turns out to be a “metaphysically impossible world”, my thought would not be a thought at all unless it conforms to logic. ... For to say that thought, in the normative sense of *judgment which is capable of truth*, necessarily conforms to logic is not to say something which a metaphysics has to *explain*. To explain anything *presupposes* logic; for Kant, logic is simply prior to all rational activity. (1994, 247)

Of course, this passage raises a lot of questions. One point is clear, though: if logic is the form of thinking, then illogical thought is not thought at all.

If this diagnosis is right, then it offers a clue for a way beyond normativism. A better philosophy of logic should try to develop the Kantian insight that the basic way logic relates to thinking is by providing its constituting structure.

0.3 Logic as constitutive of thinking

While this Kantian insight is – to my mind – a compelling one, it has been neglected in contemporary philosophy of logic. One reason for this, I think, is that the insight

³I'm not pretending to do responsible Kant scholarship here; I'm taking Kant's ideas as a clue for contemporary philosophy of logic. For further discussion of Kant's views, see Tolley (2006), Stang (2014), Merritt (2015), Lu-Adler (2017) and Boyle (2020).

has been articulated into an account whose central theses are mostly false. Rejecting these theses has led philosophers to neglect the insight that they purport to capture.

In this dissertation, I show that the insight is better articulated – with the help of some more recent philosophical material – into an account that looks very different. In this section, I’ll spell out the contrast between what the Kantian insight is commonly taken to mean and what I take it to mean. Following this, I’ll explain my usage of ‘logic’.

Any account of logic as the form of thinking needs to answer a series of difficult questions. Why does thinking need logic? Is there one specific logic that it needs? If logic is the form of thinking, does that mean it’s impossible to think illogically? Where, exactly, is the line between thinking and non-thinking? What happens when you fail to think? What is wrong with failing to think?

At a high level of abstraction, contemporary philosophers who seek to develop the Kantian insight into a more determinate theory have answered these questions as follows. (I don’t claim that each of the authors cited would accept all six theses, but each would accept most of them, and together they form a coherent picture.)

- (T1) The basic relation of logic to thought is that logic makes it possible to represent reality in a way that is either true or false. Without logic, your representations might have correctness-conditions, but not truth-conditions (Rödl 2006, 353).
- (T2) There’s one universally correct logic. Someone who followed a different logic would be a logical alien, which makes no sense (Conant 1992).
- (T3) Illogical thinking is impossible. While subjects can make logical errors, the acts in which they make those errors are not acts of thinking (Tolley 2006, 385).
- (T4) There’s a sharp line between thinking and non-thinking. At any given moment, a subject is either thinking or they aren’t; there are no in-between cases. Thinking is a distinctively self-conscious representational activity (Kimhi 2018, 53), and you either have this self-consciousness or you don’t.
- (T5) When you fail to think, nothing at all happens at the level of thought, even if you’re under the impression that you are thinking. You aren’t exercising your logical or conceptual capacities in such cases (Moore 2000, 198).
- (T6) Logic is normative purely in virtue of being constitutive of thinking. Illogical thinking is bad just because it isn’t thinking at all (Merritt 2015, 483).

By contrast, despite fully expressing the Kantian insight that logic is the structure of thinking, the account I develop in this dissertation rejects each of these theses.⁴

Let me elaborate on my account a little before stating my alternatives to T1–T6. The core claim in my account is

Logical Constitutivism A subject’s representational activity counts as thinking only if it manifests sensitivity to logical rules.

In arguing for this claim, I start from the theory of understanding – the theory of how we understand concepts and thoughts (Peacocke 2008). In my view, the right theory of understanding has to allow for the flexibility of our grasp of thoughts; I show that this flexibility requires that thinkers be sensitive to, or *follow*, logical rules. So thinking requires following logical rules.

Now, being sensitive to logical rules doesn’t mean explicitly knowing logical propositions, and nor does it mean always conforming to them. Rather, it means that we tend to find instances of those rules compelling. For example, the inference from p and *if p , then q* to q is tempting, and judging p and *not- p* seems unacceptable.

Putting this account of rule-following together with the requirement that thinkers follow logical rules generates two limits on illogical thinking. First, thinkers have to tend to find instances of logical rules compelling. Second, thinkers can’t think in obviously illogical ways. As a result, in my account, thinking has to be logical – but not perfectly so.

I develop two important consequences of this view. First, anyone who thinks that there’s a limit on illogical thinking must hold that sometimes we fail to think by virtue of illogicality. In such cases, we produce nonsense. But theorists who accept this consequence have an obligation to explain what nonsense is, and how non-thinking can seem the same (from the inside) as thinking. I show that, when we fail to think, we can exercise our conceptual capacities in a way that makes our activity seem like thinking.

Second, while an account of the normativity of logic doesn’t fall directly out of logical constitutivism, I explain how to build such an account with logical constitutivism as an integral part. By giving this account, I recapture the intuition

⁴This way of presenting my view is inspired by Paoli’s (2007) discussion of relevant logic. Tyke Nunez has also developed a modified Kantian account, but it remains committed to T1 (Nunez 2018, 1175), T2 (2018, 1176), T3 and T4 (2018, 1164), though not T5. Nunez aims to distance himself from T6 (2018, 1162), but, as I’ll argue in Chapter 6, he doesn’t go far enough.

that logic is normative without ceding ground to the normativist error that logic is fundamentally evaluative of thinking rather than constitutive of it.

As you might suspect, my answers to the questions I raised earlier are very different from T1 to T6.

- (T1') The basic relation of logic to thought is that logic makes possible the conceptual freedom that thinking involves – the freedom to question, modify or even reject seemingly obvious conceptual connections without loss of understanding.
- (T2') There isn't just one correct logic for thinking, but a range of correct logics, any one of which a thinker might follow. A logic for thinking does, however, have to contain certain minimal principles: Modus Ponens and Non-Contradiction, and possibly others. So there's room for different individuals, different communities or different species of thinkers to follow different rules, but within limits.
- (T3') Illogical thinking is possible. A thinker might accidentally judge a contradiction, or reject a direct consequence of their beliefs, because they were inattentive, because the thoughts involved were too complicated to understand, or simply because their logical capacities misfired on that occasion. However, thinking can't be illogical to an arbitrary degree.
- (T4') The line between thinking and non-thinking is vague, because it is vague how often your capacities can misfire before you fail to count as being sensitive to logical rules at all. As a result, there might be borderline cases between thinking and non-thinking.
- (T5') When you fail to think, you can exercise your logical and conceptual capacities within the scope of a pretence that you're thinking. This allows you to reason with nonsense and embed it within that-clauses.
- (T6') Being constitutive alone is not enough for normativity: many activities have internal standards without those standards being normative. Logic is normative because thinking is necessary for human flourishing.

These claims are defended in Chapters 1 to 6, respectively.

0.3.1 What I mean by 'logic'

You might have noticed that I haven't said what I mean by 'logic'. In this section, I'll rectify that omission and clarify the order of conceptual priority in my discussion.

Before this, a brief note on symbolism. Except where otherwise stated, lower-case italic letters like p and q range over thoughts; upper-case italic letters like P and Q range over sentences in artificial object-languages; and capital Greek letters like Γ and Δ range over sets of either thoughts or sentences, depending on context. Instead of using corner quotes, logical symbols like \rightarrow are used as names of the identical object-language symbols; concatenation in the metalanguage (i.e. in the language I'm using now) stands for concatenation of the object-language symbols (Church 1956, §8).

In this dissertation, I'll use 'logical concepts' to mean a set of concepts which express topic-neutral rational relations among thoughts, and which are constituted by rules of inference, or 'logical rules'. I'll use 'logic' more loosely to refer either to logical concepts or logical rules. Logic is topic-neutral in the following sense: if an inference from some premises Γ to a conclusion q is logically valid, it is so regardless of the subject-matter of the thoughts in Γ and q . Here's an example of a logical rule that we'll return to in later chapters.

Modus Ponens From p and *if p , then q* , infer q .

It's natural to write down logical rules as imperatives, but this notational choice isn't meant to beg any philosophical questions: in particular, it isn't meant to suggest that the rule on its own has normative force.⁵ For present purposes, all we need is a way of specifying the class of transitions that the rule licenses.

Of course, this isn't the only thing one might mean by 'logic'. For example, it's far removed from a conception of logic as a science continuous with other sciences (Williamson 2013). But the idea that logic is a set of concepts which express topic-neutral rational relations among thoughts, constituted by rules which all thinkers must follow, is in line with a long tradition. It goes back at least as far as Kant, who wrote that 'the universal and necessary rules of thought in general can concern merely its *form*' (*Jäsche Logic* 528), and Hegel, who wrote that logic's content 'is nothing but our own thinking and its ordinary determinations' (*Encyclopedia Logic* 45).

This conception was taken up and reworked by early analytic philosophers like Frege, who wrote that logic's task is 'saying what holds with the utmost generality for all thinking, whatever its subject-matter' ('Logic' 128), and Wittgenstein, who wrote that 'the sole logical constant [is] what *all* propositions, by their nature, [have] in

⁵For this point, see Harman (1986, 5).

common with one another' (*Tractatus Logico-Philosophicus* 5.47), and that thinking illogically is impossible (5.4731). I think these philosophers were on to something, and what I'm doing is continuous with their work – although I differ from them on many points of detail, just as they differed from each other.

So much about my conception of logic. But I want to be clear about the role this conception plays in my discussion. I'm not assuming that thinkers must possess concepts which express topic-neutral rational relations among thoughts. Nor am I assuming that the concepts which play this role are ones which we would ordinarily think of as logical – concepts like negation or the conditional. Rather, both of these are claims that I will argue for in the course of the discussion. So you don't have to agree with these claims at the outset: this is just a preview of where we're headed.

At this point, you might worry that this order of priorities brings me dangerously close to psychologism – the view that logical rules are either generalizations about how we in fact think, or natural laws of thinking (Husserl, *Logical Investigations* §21-24). The short answer is that it doesn't: my claim is that logic is constitutive of thinking, not descriptive of it. The longer answer is that this worry rests on conflating the order of knowledge with the order of being. My argument starts from the activity of thinking, and shows that a certain relation to certain rules is a condition of the possibility of that activity. (Roughly speaking, it's a transcendental argument.) Then, I follow Kant *et. al.* in calling rules that play this role 'logical'. So in the order of knowledge, I begin from thinking and end with logic.

But it doesn't follow from this that logic is grounded in thinking. Just the opposite: logic is part of what makes thinking possible. In the order of being, thinking is grounded in logic. If q is a logical truth, it is true regardless of its relation to thinking, and if there were no thinkers around, q would still be true. So I wholly agree with Frege that you can only explain why a logical truth is true by reducing it to another logical truth (*Basic Laws* xvii), and not by appealing to any facts about thinking. As Woleński (2003, 186) nicely puts it, if q is a logical truth and Δ is a set of psychological truths, you only have $\Delta \vdash q$ if you have $\vdash q$; the psychological truths are never needed as premises.

Getting these priorities straight matters because it bears on the connection between my account and the *deductive systems* produced by mathematical logicians – that is, theories about consequence relations that hold among sentences of artificial languages. My conception of logic didn't make any reference to deductive systems. In fact, there's no upfront guarantee that logic and deductive systems have anything to

do with each other. This means that my account is upstream of several contemporary debates: I don't presuppose a criterion for demarcating logical from non-logical constants (Tarski 1966; Etchemendy 1990) or a theory of logical necessity (McFetridge 1990). Nor do I have to start with any assumptions about whether there's one correct logic (understood as a deductive system) or many (Beall & Restall 2005).

This isn't to say that my account has no bearing on these debates – only that its bearing will depend on what the connection is between logic and deductive systems, and this connection needs to be argued for, not assumed. Moreover, any connection that does exist between logic and deductive systems is likely to be weaker than identity, so not everything that's true of one has to be true of the other. If logic is constitutive of thinking, it doesn't follow that the Hilbert-style axiomatization of classical logic you teach your students is constitutive of thinking.

Of course, it *is* an interesting question what the connection is between logic in my sense and the theories produced by logicians: are the rules we think with the same as the rules in some formulation of, say, classical or intuitionistic logic? Is there just one set of correct rules, or are there many different ones? I investigate these questions in Chapter 2. It would also be interesting to explore whether my account generates a demarcation criterion, or an explanation of logical necessity, but I haven't done that here.

0.4 Dissertation outline

The next four chapters comprise **Part I** of the dissertation. These chapters develop logical constitutivism.

In **Chapter 1**, I argue that thinkers have to follow logical rules. The argument begins with a holistic account of understanding, which recognizes that thinkers are free to question, modify or even reject seemingly obvious conceptual connections without loss of understanding. Much recent work in the theory of understanding has emphasized this kind of freedom and its importance in the development of thought. I show that, on this holistic account, thinkers have to have general abilities to respond to the rational relations among their thoughts. These abilities are constituted by possessing logical concepts. So thinkers have to possess logical concepts, and follow the rules that govern those concepts.

In **Chapter 2**, I explore the interactions between logical constitutivism, disagreement about which logic is correct and logical pluralism. I argue that my view

doesn't rule out the possibility of rational disagreement about which logic is correct and that it allows for two kinds of logical pluralism: different logics might be correct in different domains, and thinkers might also use different logics to think about the same domain. In this sense, there are a range of correct logics, any one of which a thinker might follow. I briefly comment on which logics meet the requirements.

In **Chapter 3**, I consider what it is to follow logical rules. I argue that we follow logical rules by exercising logical capacities, which display a distinctive first-person/third-person asymmetry: a subject can find the instances of a rule compelling without seeing them *as* instances of a rule. Recognizing this asymmetry helps to resolve certain puzzles about the sense in which reasoning involves following rules. On the resulting view, we can have and exercise logical capacities without always conforming to the rules.

In **Chapter 4**, I argue for two limits on illogical thinking. First, thinkers have to *tend to* be sensitive to logical rules. In response to an objection, I observe that this limit is vague, and show how different treatments of vagueness can be brought to bear on the problem. Second, I argue, thinkers can't think in an *obviously* illogical way. I clarify the sense in which it's impossible to judge a contradiction and show that this is consistent with *reductio* proofs.

That concludes my development of logical constitutivism. The chapters in **Part II**, which were written as standalone papers, spell out two important consequences of the account.

My view entails that sometimes, because of illogicality, we fail to think. In such cases we produce nonsense. Anyone who accepts this consequence has an obligation to explain what nonsense is, and how failing to think can seem the same (from the inside) as thinking. So in **Chapter 5**, I give an account of nonsense. The account I give doesn't depend on the views defended elsewhere in the dissertation, and applies to a broader class of failures to think than those due to illogicality alone: for example, if category mistakes are nonsense, then my account applies to them too. I argue that an adequate account of nonsense has to meet two constraints: it has to explain how our engagement with nonsense is like our engagement with sense, but it has to avoid assimilating nonsense to sense. I show that some existing accounts fail to meet one constraint or the other, and propose an account that meets both. On my proposed account, our engagement with nonsense is an exercise of imagination that draws on the role of logical form in understanding.

You might wonder whether logic is normative on my account. **Chapter 6**

addresses this question. I show that, while the normativity of logic doesn't fall directly out of logical constitutivism, it's possible to build an attractive account of logical normativity which has logical constitutivism as an integral part. I start by trying to develop an account of logical normativity on analogy with constitutivism in metaethics, and show that this attempt fails: even if logic divides thinking from non-thinking, this doesn't show that we have any reason to think. However, logical normativity does follow if we supplement my account with a claim about the value of thinking. In particular, I argue that thinking is necessary for human flourishing, and that this is the source of the normativity of logic. The resulting account also deals nicely with several vexing problems in the literature on the normativity of logic.

Part I

Developing logical constitutivism

1 Why thinking needs logic

1.1 Introduction

In this chapter, I argue for a precise version of the thesis that logic is necessary for thinking. The idea that there is something essentially logical about thinking, that thinking cannot take place outside of a logical framework, is an old one in analytic philosophy. It can be found in early works of mathematical logic, like Boole’s *The Laws of Thought*, which takes logical laws to govern the basic operations of the mind, such that mathematical logic is a clue to the ‘Constitution of the Intellect’, and Frege’s *Begriffsschrift*, which presents its symbolism as a ‘Formula Language for Pure Thought’. The same idea shows up in a strong form in the early Wittgenstein’s insistence that thought ‘can never be of anything illogical’ (*Tractatus Logico-Philosophicus* 3.03), i.e. that it is ‘impossible to represent in language anything that “contradicts logic”’ (3.032) and more weakly in Quine’s dismissal of the possibility of ‘prelogical mentality’ (1960, 58; 1974, 80).

But the idea is not quite so popular these days: it is sometimes explicitly denied, for example by Brandom (1994, 383), and is certainly not taken for granted. One reason for this is that the rather rigid theories of understanding that motivated the idea in early analytic philosophy are themselves less popular than they were.¹ This makes it less obvious, on contemporary views, why thinking would need logic. In this chapter, I show that even in a contemporary theory of understanding which emphasizes the flexibility of our grasp of concepts, thinkers must follow logical rules – indeed, that logic makes this flexibility possible.

1.1.1 Terminology

Before sketching my argument, let me clarify some terms. I’ve been talking about ‘thinking’. This word is used in various ways. Sometimes it is used as a catch-all for activity with representational content – as, for example, by Descartes in the Second Meditation: ‘Well, then, what am I? A thing that thinks. What is that? A thing that

¹Another likely culprit is that the idea gets conflated with psychologism. Absent this conflation, the idea is not directly threatened by the deluge of psychological results showing that our beliefs are sometimes irrational (see Bergamaschi Ganapini (2019, 4) for a summary and references).

doubts, understands, affirms, denies, wants, refuses, and also imagines and senses.’ I’m using the term in a more discriminating sense. ‘Thinking’ is not a term for the genus, representational activity, but for a particular species of representational activity. It is the representational activity composed of acts of judging and inferring, acts whose contents are propositions or ‘thoughts’. Exactly what distinguishes thinking from other kinds of representational activity is a difficult question, and I’m not proposing to answer it here. I am going to presuppose, however, that thinking some thought p requires understanding p ; this link may not hold for other kinds of representational activity, such as asserting.

A couple more qualifications will help to focus my discussion. First, in a full account of thinking, we might also want to account for peripheral cases such as entertaining a thought or reasoning under a supposition. Here I focus on the core cases. Entertaining and supposing raise different issues, because some of the norms of judging and inferring seem to be suspended in these cases: we have good reason not to judge a contradiction, but no reason not to entertain one. Second, I’m assuming that judging is binary rather than allowing for varying degrees of confidence. If we instead conceived of judging in terms of degrees, we might hold that the axioms of probability have a role similar to that which I defend for logic. This suggestion could find additional support in the claim that the axioms of probability are merely ‘a way of applying standard logic to beliefs, when beliefs are seen as graded’ (Christensen 2004, 15). But without a detailed consideration of whether logic and probability are so closely related, I prefer to leave this issue open.

Next, ‘logic’. Many people use ‘logic’ to refer to a theory of the sort produced by logicians, or the discipline to which those theories belong. Typically, these theories are about a consequence relation that holds among sentences of a formal language. That is not what I will mean. I use ‘logical concepts’ to mean a set of concepts which express topic-neutral rational relations among thoughts, and which are constituted by rules of inference, or ‘logical rules’. An example of a logical rule is Modus Ponens:

Modus Ponens From p and *if p , then q* , infer q .

My claim is that thinkers have to possess logical concepts, which they do by following logical rules. But it’s important to be clear on the conceptual order here. I’m not presupposing that thinkers have such concepts, or that these concepts are ones we would ordinarily think of as logical. Rather, I’m going to *argue* that thinkers must have such concepts, and that they include at least some of the concepts we would

ordinarily think of as logical – namely, the conditional and negation. The premises of my argument don't presuppose anything about logic: in fact, the word 'logic' doesn't appear at all in sections 1.2 or 1.3.

1.1.2 Thesis and argument

My thesis is that a subject's representational activity counts as thinking only if it manifests sensitivity to the rules which constitute the concepts of the conditional and negation. More concisely, I argue for

Logical Concepts A thinker must possess concepts of the conditional and negation.

It follows from my discussion of these concepts that possessing them involves sensitivity to logical rules. In this chapter, I'm going to rely on our intuitive sense of what it is to 'be sensitive to' or 'follow' a rule; in Chapter 3, I address this question head-on. For now, the rough idea is that you can follow a rule without always conforming to it, but you have to tend to conform, and to conform because that is what the rule says, not just by accident (Wedgwood 2006).

By way of initial motivation for this thesis, observe that there seems to be a limit on how incoherent someone's representational activity can be while they still count as thinking. As Jane Heal puts it, '*completely* unsuccessful and chaotic thought, thought in which no shred of truth or rational connectedness is discernible, is an incoherent notion' – although of course this 'does not deny the possibility of extremely bizarre beliefs, great degrees of muddle, contradiction and so forth' (1989, 89).

My argument for this thesis goes in three main steps. First, in 1.2 I develop a holistic constraint on understanding a thought. I start with the idea that understanding any thought requires appreciating its rational relations to some other thoughts. I use some cases of 'deviant understanding' to motivate a holist reading of this idea as against a more rigid inferentialist reading of it. Second, in 1.3 I give an explanation of why this holism holds: namely, that thinkers must have general abilities to respond to relations of support and tension among thoughts. I show that this explanation fits the phenomenon better than its competitors. Third, in 1.4 I argue that we have these abilities by possessing logical concepts – in particular, the conditional and negation. Following this, in 1.5 I show that this view makes sense given the Quinean roots of holism about understanding and in 1.6 I conclude.

1.2 The holistic constraint on understanding

You're doodling on a piece of paper. What does it take for a mark on the paper to represent an eye?² A mark on its own isn't naturally interpreted as an eye or anything else. One way of getting a mark to represent an eye is by putting it in the context of a face: for example, a dot can represent an eye if you put another dot next to it, and draw a circle around them, and maybe a line underneath the dots to make things extra-clear. Another way is to make the mark look like an eye: in other words, to put it in the context of other marks which represent parts of an eye – pupil, iris, eyelashes. And there are other ways.³ While there's no single context you need to make a mark represent an eye, you do need some context or other.

In this section I'll defend a holistic constraint on understanding a thought. It says that understanding is, in this way, like drawing an eye.

Holism Understanding a thought p requires sensitivity to some of its rational relations to other thoughts.

I'll say more about technical terms like 'sensitivity' and 'rational relations' below; for now, the rough idea is that understanding p involves relating p to other thoughts. While there is no single set of relations you need to grasp in order to understand p , you do need to grasp some set or other. I'm not proposing this as a full account of understanding, but as a necessary condition for understanding.

Before arguing for holism, I want to distinguish it from two other views with which it might be confused. First, holism about understanding is different from holism about meaning (Heal 1994). Holism about meaning says that the meaning of a sentence s is constituted by the meanings of all the sentences in the language to which that sentence belongs, such that if there is a change in the meaning of any other sentence, then the meaning of s changes (Dummett 1975). This has the unpalatable consequence that, if two speakers have slightly different idiolects, then they can't mean the same thing by any sentence. At the level of language, holism about understanding does entail that understanding s requires understanding some range of other sentences, but it doesn't follow that the meaning of s depends on the meanings of these other sentences. Holism about understanding is more closely

²This example is from Heal (2017, 312).

³They don't all involve more drawing. You might just declare, 'this is an eye.' But then the utterance provides context for the mark; indeed, the meaning of the mark is parasitic on the meaning of the utterance.

related to the metasemantic thesis that conferring a meaning on s requires conferring a meaning on some range of other sentences. It's consistent with this that the meaning of s remains the same while the meaning of some other sentence changes.

Second, holism about understanding is different from holism about confirmation, Quine's (1951) view that it doesn't make sense to speak of confirming or falsifying a single statement on its own, as only a collection of statements can be confirmed or falsified. Holism about confirmation might be false while it remained true that understanding a thought requires placing it among a range of other thoughts. Again, then, the two theses are distinct.⁴

I'll argue for holism as follows. First, I'll argue for the general claim that understanding one thing requires understanding others: in short, *understanding requires a setting* (Heal 1994, 334). Second, I'll motivate holism in particular by reflecting on cases of *deviant understanding*. These cases were first introduced by Tyler Burge and taken up by Timothy Williamson as counterexamples to purportedly analytic truths and analytic inferences, but they are also examples of a distinctive kind of conceptual freedom which provides support for holism.

1.2.1 Understanding requires a setting

In this section I argue that understanding requires a setting: understanding a thought p requires sensitivity to its rational relations to a range of other thoughts.

By 'rational relations', I mean at least the following two kinds of relations: relations of rational support, where q is a consequence of p , and relations of rational tension, where r is incompatible with p . By 'sensitivity', I mean that the subject has to grasp, or recognize, the way that taking up or revising a position on q or r could give them reason to take up or revise their position on p . This characterization is intentionally loose, and I'll return to it below. For now, I'm leaving it open whether sensitivity involves knowledge, or even beliefs, about these rational relations.

Let me clarify a few points. First, these relations need not be formal ones, like

⁴Quine suggests that the two theses go together (1974, 38): 'An observation may refute some chunk of theory comprising a cluster of sentences, and still leave us free to choose which of the component sentences to continue to count as true and which to abandon. ... The semantical relation of observation to the theoretical language is similarly intricate and indirect, since we learn the language only partly by associating terms or sentences directly with observation, and partly by linking them to one another. The evidence relation, in all its intricacy, and the semantical relation, in all its intricacy, are coextensive still.' But this coextensiveness depends on Quine's account of understanding; in principle, you could hold Quinean views about evidence but not about understanding.

the tension between p and $not-p$ – indeed, it’s a mistake to hold that all rational relations are formal (Brandom 1994, 97ff.). Second, in a fuller treatment we would also have to include cases where p and r together support q , although neither p nor r support q on their own: where the support-relation is not a V-shape with separate lines from p and r to q , but a Y-shape (Berker 2015, 330). Similarly, we would have to include cases where p , q and r are in tension, although no two of them are in tension. However, these cases are more demanding than the binary case of support or tension between two thoughts, so I leave them aside. If the binary case already requires logic, this conclusion won’t be threatened by adding the higher-arity cases.

Third, there might be weak relations of support where p increases the likelihood of q by some small degree, but not enough to license believing q on the basis of p (or the converse for tension). Accounting for such cases would require me to consider degrees of belief and the way the axioms of probability govern credences, and I’m not going to do that here. Still, I take it that a thinker has to be sensitive to *some* rational relations which are strong enough to justify an inference, and indeed that thinkers do sometimes correctly infer on the basis of such rational relations. Ichikawa & Jarvis (2013) also develop a view in which understanding a thought involves sensitivity to its objective rational relations, but they take these relations to be *conclusive* – indefeasible except in cases of compromised rationality (2013, 44). I’m not going to assume that the rational relations I consider are conclusive, but I’ll comment in 1.4.1 on what difference this further assumption would make.

Finally, while my discussion focuses on relations of support and tension between thoughts, and while I take sensitivity to such relations between thoughts to be essential for thinking, it’s important to note that support and tension don’t only hold between thoughts. There are also the relations that Sellars called ‘language-entry’ and ‘language-exit’ (1974, 423). A thought can be rationally supported by perceptual evidence, and perhaps by other kinds of evidence, such as testimony; by the same token, it can stand in tension with evidence. A thought can rationally support actions, as it does in belief-desire explanation, and by the same token can stand in tension with them. I think these connections with evidence and action are necessary for thoughts to have content. This point goes back to Kant’s claim that ‘[w]ithout sensibility no object would be given to us’ (*CPR* A51/B75) and Burge’s (1977) argument that having beliefs *de dicto* presupposes having beliefs *de re*. But entry and exit aren’t sufficient; grasp of relations among thoughts is necessary too.

The thought that understanding requires a setting is suggested by an observation that Gareth Evans made in arguing against simple dispositional accounts of belief. Building on Ryle's observation that intelligent capacities are 'not single-track dispositions, but dispositions the exercises of which are indefinitely heterogeneous' (1949, 44), Evans noted a difference between the 'beliefs' we attribute to animals and those of human beings (1982, 336-7).

The rat manifests the 'belief' [that something is poisonous] in only one way – by not eating – whereas there is no limit to the ways in which the ordinary belief that something is poisonous might be manifested. The subject might manifest it by, for example, preventing someone else from eating the food, or by giving it to a hated enemy, or by committing suicide with it. ... One who possesses a belief will typically be sensitive to a wide variety of ways in which it can be established (what it can be inferred from), and a wide variety of different ways in which it can be used (what can be inferred from it).

The rat can only do one thing with its 'belief', while the human being can do an open-ended range of things with it, including being able to support the belief in various ways and derive various consequences from it. This suggests that sensitivity to some of the rational relations of a belief is essential to having the belief in the first place. (This is not to say the rat has no representational content at all – just that it doesn't have beliefs.)

Consider what would follow if we denied that understanding requires a setting. It follows, first of all, that someone could think p while being wholly insensitive to the rational relations between p and other thoughts. But then it would seem that any other thoughts the thinker understood would be making no contribution to the thinker's understanding of p . So it would seem possible that someone could think a single thought p as the entirety of their mental life.⁵ Now, I don't have much of an argument against this possibility, and it has occasionally been accepted, for instance by Jerry Fodor (1987, 89). But to me it seems absurd. I can't imagine a case where we would have reason to attribute one thought to a creature without having reason to attribute any others.

Of course, without accepting interpretivism – that to have a thought is to be interpretable as having it – it doesn't follow that there could be no such creature. But

⁵Kwong (2007) denies that possessing a concept requires possessing others, but maintains that *acquiring* a concept requires possessing others. But on Kwong's view the one-thought mind is still metaphysically possible – its concepts just have to be innate rather than acquired.

the fact that we are pressed to acknowledge a possibility which we could never have reason to take to be actual still seems problematic. What would make it the case that the creature is thinking p rather than some other nearby thought p' , if the creature isn't able to recognize any of the grounds or consequences that distinguish p and p' ? Note that, while some have argued that our thoughts are perfectly determinate (Ross 1992; BonJour (ms.)), I'm not taking such a strong position here. Quine (1960, §15) may be right that all our thoughts are indeterminate to some degree. It doesn't follow that *any* degree of indeterminacy is consistent with thinking; it seems more likely that there's a limit on how indeterminate a thought can be, and the creature who can only think p is past that limit.

Suppose, for example, that p is the thought *this is square* (had while the creature is looking at a table). What would make it the case that our creature, incapable of having any other thoughts, is thinking *this is square* rather than, say, *this is rectangular*? In any situation where the former is true, the latter is, too. The ordinary way of telling these cases apart rests on the fact that *this is square* requires more specific grounds than *this is rectangular*, and that that it has broader consequences – for instance, it lets you infer *all of its sides are the same length*. But this sort of thing is, by hypothesis, unavailable. Nor can we appeal to the fact that the table the creature is looking at is square: this fact is only there to appeal to in cases where the creature is thinking truly, but it's surely possible to think falsely that something is square (Millikan 1984, 7). In any event, if the table is square, it's rectangular too, so this consideration fails to distinguish the two thoughts. Indeed, the table also has a colour (say, green) which is also visible to the creature, so this consideration fails even to distinguish *this is square* from *this is green*. It seems to me that if we have grounds to attribute the thought *this is square* but not *this is rectangular* or *this is green*, then we must also have grounds to attribute some other thoughts.

You might suggest that language-entry and language-exit are sufficient for understanding: in other words, although the creature doesn't recognize any links between p and other thoughts, it's sufficient if it recognizes links between p , evidence for p and action supported by p . But there are two problems with this suggestion. First, while they are likely essential for contentful thought, links with evidence and action on their own do not confer determinacy. The rat might be disposed to form a representation in response to the colour or smell of its food, and it might be disposed not to eat the food when it has formed this representation, but that isn't enough to attribute to it a determinate thought: a thought is 'conceptually too rich for the

purpose' (Dummett 1994, 122). In the case of human beings, links with action and evidence do seem to justify attributing determinate thoughts, but this is because they manifest connections with other thoughts as well: if the subject gives the food to a hated enemy, this might justify attributing the belief that it was poisoned, but only against a background of other beliefs.

Second, a subject whose representations are tied this closely to evidence and action violates a requirement on thought which Elizabeth Camp (2009, 288) calls 'stimulus-independence':

genuine thought involves a clear distinction between representation and represented, so that the former can occur even in the latter's absence. ... Because thoughts are at least partly constituted by their contents, understanding a thought requires grasping the conditions required for its satisfaction. But if a thinker really does grasp those conditions of satisfaction, as opposed to simply being confronted by the conditions themselves, then its grasp of those conditions should be relatively independent of its current circumstances. Otherwise, the world, and not the thinker, is shouldering the bulk of the representational burden. And if this is so, then that "thinker" really is just a passive reactor.

Camp's idea – which, as she notes, has roots in Dummett (1994, 123) and McDowell (1994, 57) – is that understanding a thought requires an ability to think it independently of being confronted by the situation the thought is about. But a creature which only makes language-entry and language-exit transitions, and not transitions between thoughts, is unable to think a thought without being confronted by, or taking itself to be confronted by, evidence for that thought, so it lacks stimulus-independence. Without discussing the connection between stimulus-independence and understanding, I don't want to put too much weight on this point. But to the extent that the connection is plausible, it tells against the suggestion that language-entry and language-exit are sufficient for understanding.

So there's strong reason to accept that understanding requires a setting. As it stands, however, there are different versions of this claim, depending which sets of rational relations are said to be required for understanding. One version is inferentialism – the view that understanding a thought requires recognizing a particular set of rational relations involving that thought (Peacocke 1992). But it's possible to maintain that understanding requires a setting while refusing the inferentialist gloss, which involves drawing a distinction between 'canonical' and 'non-canonical' rational relations – a rational relation being canonical for p iff recognizing

that relation is necessary for understanding p . Holism accepts that understanding requires a setting without commitment to this distinction.⁶ In the next section, I introduce cases of deviant understanding, which support holism specifically.

1.2.2 Deviant understanding

In this section I'll discuss cases of deviant understanding – that is, cases where a subject understands a thought p in a deviant way, rejecting obvious truths containing p or obviously valid inferences involving p . I'm not going to define deviance or obviousness, as my argument doesn't require any precise distinction between deviant and non-deviant understanding. The cases I'll discuss were introduced by Burge, and taken up by Williamson, as counterexamples to purportedly analytic truths or inferences, but my interest here is in the distinctive freedom that subjects in these cases display. I'll argue that making room for this phenomenon requires holism.

Consider the thought *this is a sofa*. Most people consider it obvious that *this is a sofa* entails *this is for sitting on*; it might seem that anyone who fails to accept this consequence fails to understand *this is a sofa*. But Burge (1986) shows that, with a bit of imagination, we can tell a story in which a speaker understands *this is a sofa* but rejects the link in question. Suppose we meet somebody who is able to identify sofas and uses the word in a syntactically appropriate way, but who staunchly denies that they are for sitting on. When we ask why not, we get the following response:

“Sofas are, in fact, items of religious worship. That's why we traditionally keep them in the most central part of the house, and why so much attention is lavished on their design and materials. Of course, most people don't realize this; they think that a sofa is just a regular piece of household furniture. What people don't know is that, in the early years of the 20th century, there was an extensive – and highly successful – propaganda campaign to eradicate the sofa cult. And what better way to degrade a religious item than to convince people everywhere to sit on it? As I say, the campaign was so successful that, today, hardly anybody knows about it: but none of this changes the fact that sofas are really not for sitting on...”

And so on in this rather tiresome vein. What should we say about this? Burge's view – which I think is the right one – is that this person understands the thought

⁶There are strong arguments that in *some* cases some rational relations are canonical (Balcerak Jackson 2009; Boghossian 2012), but I doubt this holds in every case.

this is a sofa, even though they reject the obvious entailment. They understand the thought in a deviant way.

Note that understanding the thought *this is a sofa*, and in particular, understanding the concept *sofa*, amount to more than just referring to sofas. Merely being able to pick out the same class of objects that we pick out using the concept *sofa*, and having a term for objects in that class, is enough to count as referring to sofas; no overlap with our own conceptual connections is required. Burge isn't making the weak claim that his character's use of the word 'sofa' refers to the same things that we refer to; he's making the strong claim that the character's use of the word 'sofa' expresses the same concept that we express, such that he can reject the same thought that we accept.

Williamson (2006; 2007, ch. 4) has developed a range of similar cases, focusing on purportedly analytic statements – that is, truths the acceptance of which is a necessary condition on understanding them. Consider Williamson's counterexample to the purportedly analytic statement 'Every vixen is a vixen'. Williamson's character 'Peter' thinks that universal generalizations presuppose the existence of an instance, so 'Every vixen is a vixen' presupposes that there are vixens. Peter also believes that there have never been any vixens (2006, 9-10):

For he spends far too much time surfing the Internet, and once came across a site devoted to propagating the view that there are no foxes, and therefore no vixens, and never have been: all the apparent evidence to the contrary has been planted by a secret international agency; for sinister purposes best known to itself, it produces elaborate fox-hallucinations. Being a sucker for conspiracy theories, Peter accepted this one.

As a result, Peter thinks 'Every vixen is a vixen' is false. But we're still inclined to say he understands it.

The subjects in these deviant cases display, to an extreme degree, a freedom which, in less extreme forms, is a familiar part of our cognitive lives. It's possible to suspend commitment to, or even reject, seemingly obvious beliefs or inferences without loss of understanding; in particular, it's possible to use a concept in a way that belies some apparently definitional link to other concepts while still counting as using that concept. We have the conceptual freedom to question or modify some of the rules that constitute our inherited representational system.

This phenomenon goes beyond the rather fruitless eccentricities described in the cases above. It doesn't merely make possible counterexamples to analyticity, but also

plays a role in the development of thought in general (Williamson 2007, 126; Rattan & Wikforss 2017, 283). Here are two examples. First, scientific progress sometimes involves rejecting connections that at one time would have seemed definitional, or at least impossible to doubt. For example, in Newtonian mechanics it would have seemed impossible to doubt that the mass of an object is the same in all reference frames, but this assumption is given up in relativity theory (subject to some complications about the reference of ‘mass’ in the two theories (Field 1973, 467)). Second, certain sorts of social and political change are also bound up with exercises of conceptual freedom – for instance, in rejecting the idea that a marriage has to be between a man and a woman, or in changing understandings of the latter concepts (Haslanger 2000; Jenkins 2016).

The role of conceptual freedom in the development of thought in general makes it harder to dismiss than it otherwise would be. As Davidson writes in a related context, ‘A theory that could not explain irrationality would be one that also could not explain our salutary efforts, and occasional successes, at self-criticism and self-improvement’ (2004, 187). An account of understanding that had no room for deviant understanding would also have no room for the use of conceptual freedom to criticize and improve our system of thoughts. So we should make room for this phenomenon in our account of understanding.

This gives us reason to accept holism. In fact, two reasons. Negatively, it’s clear that the inferentialist version of the claim that understanding requires a setting would be inconsistent with conceptual freedom: if inferentialists were right, it would be impossible to reject a canonical rational relation without loss of understanding. So we should accept the holist version of the claim instead.

But I think conceptual freedom also provides more positive support for holism. We can see this by further reflection on the *sofa* case. The intelligibility of the subject who denies that sofas are for sitting on depends on their other beliefs (cf. Reimer 2012, 255). Some of these are orthodox beliefs about sofas themselves, e.g. that sofas are human-made artefacts rather than, say, vegetables; that they generally stay where they are, rather than moving about on their own. If the subject believed not only that sofas aren’t for sitting on, but also that a sofa is a sort of pumpkin which travels about on its own, in pursuit of its own purposes, we might be less inclined to agree that they still understood *this is a sofa*. Intuitively, the thought *this is a sofa* is held in place by a network of links to other thoughts. While the subject has cut an important link – the connection to *this is for sitting on* – enough of the other

links are still intact to keep the thought in place.

Others among the subject's beliefs play a different role. They rationalize the subject's rejection of the 'canonical' entailment. These are the beliefs about the propaganda campaign and the religious nature of sofas. While these beliefs are unorthodox, we can see how someone with such beliefs would deny that sofas are for sitting on; without these beliefs, the subject's denial would be much less intelligible.

Both sets of beliefs are essential to deviant understanding. This suggests that, 'unless a plausible background story emerges such that the "unorthodoxy is compensated for by orthodoxies at other points", the fact that [the subject denies the analytic statement] does entail failure of understanding' (Wikforss 2009, 12). In other words, a subject counts as understanding p only if there is some 'plausible background story' – some set of links to other thoughts – which underwrites the subject's understanding of p .⁷ But as the cases show, there's no single background story that is required, but many background stories each of which would be sufficient for understanding. Requiring a single background story for understanding p would leave no room for conceptual freedom involving p . And the claim that understanding p requires a background story, but no particular background story, is equivalent to holism. So cases of deviant understanding support holism.

1.3 General rational abilities

In the previous section, I began by motivating the thesis that understanding requires a setting. I then introduced cases of deviant understanding, which, I argued, show that we should accept holism as a particular version of the thesis. In this section, I ask what explains the truth of holism, and argue that it reflects the thinker's sensitivity to rational relations among their thoughts. In particular, I will suggest that the best explanation for why thinkers in deviant cases count as understanding is that the thinkers manifest general abilities to recognize and respond to relations of support and tension among their thoughts (cf. Boyle 2009, 150-151). These are

⁷Williamson implies that the role of the background story is merely practical (2006, 36): 'Although disagreement is naturally easier to negotiate and usually more fruitful against a background of extensive agreement, it does not follow that any particular agreement is needed for disagreement to be expressed in given words. A practical constraint on useful communication should not be confused with a necessary condition for literal understanding.' But without the background beliefs, why accept that the deviant speaker still understands the thought? Moreover, in denying that 'any particular agreement is needed', Williamson rules out only the inferentialist construal of the point, leaving room for the holist position that some agreement or other is, in fact, needed.

deep waters, and I don't have a proof that such abilities are the right explanation. But I will suggest that this is the feature of the cases which drives our conclusion that the deviant thinkers understand what they are saying. I'll begin by ruling out the two most plausible competitors, before spelling out the explanation I prefer.

1.3.1 Two unsuccessful explanations of holism

Why is it possible to understand p while rejecting some obvious inference involving p ? I'm going to consider two explanations in this section and argue that neither of them succeeds. While we're looking for an explanation of the truth of holism generally, I'll focus on cases of deviant understanding, as these are where holism comes apart most clearly from other theories.

The first possibility is that holism is underwritten by *social externalism*. According to social externalism, a subject counts as understanding p if, in their use of p , they intend to defer to the way p is used in their community, or to the way p is used by relevant experts (Putnam 1975; Burge 1979). For example, I count as understanding the concept *chrysanthemum* even though I'm unable to recognize chrysanthemums, and even though the inferential role I associate with the concept is indistinguishable from the role I associate with other concepts of flowering plants, like *rhododendron*, because when I use *chrysanthemum* I intend to defer to whatever 'chrysanthemum' is taken to mean by gardeners or botanists. In the present context, the idea would be that the deviant user can still intend to defer to whatever the community means by 'sofa', and so still count as understanding it.

I should note that this probably isn't the right reading of either Burge or Williamson. Burge is clear that his arguments about *sofa* take a different route from his arguments on social externalism; he distinguishes the notion of cognitive value, which he takes to be at play in the *sofa* case, from the notion of conventional linguistic meaning (1977, s. IV; 2007, s. V). In particular, Burge holds that since 'communally accepted characterizations as well as expert opinion can be doubted', the standards of correctness for using a cognitive value, like *sofa*, aren't determined by the usage 'of any person or social group' (1986, 720). Since cognitive value isn't determined by social patterns of usage, our grasp of cognitive value cannot consist in deference to such patterns.

Williamson is less clear about what drives his cases, and he does refer to Putnam's hypothesis of the division of linguistic labour and 'the way in which individual speakers defer to the linguistic community as a whole' (2006, 36). He also notes that Peter

intends for his words ‘to be understood as words of our common language, with their standard English senses’ (2006, 12). However, the structure of Williamson’s cases precludes any straightforward explanation in terms of social externalism, for the simple reason that the deviant speakers Williamson describes are experts (Boghossian 2012; Murzi & Steinberger 2017). Peter is said to ‘have published widely read articles on the [semantic] issues in leading refereed journals of philosophy, in English’. For a social externalist, experts are the people non-experts defer to; experts themselves don’t defer. Peter’s understanding of ‘Every vixen is a vixen’ can’t consist in his deference to other speakers or to the speech community as a whole.

Of course, we can ask whether social externalism is the right explanation for the cases regardless. But there’s good reason not to invoke it here. It fails to explain the importance of rationality in the cases – the feature that Wikforss calls a ‘plausible background story’. The sofa-worshipper’s denial of an obvious truth is rationalized by his other beliefs about sofas – the story about the propaganda campaign and religious worship. If social externalism were driving the case, the background story would be beside the point. It would be sufficient to say: one day somebody believed that sofas are not for sitting on, while intending to defer to their speech community’s use of ‘sofa’, and still understanding *sofa* in virtue of this deference. The fact that this bare case is wholly unconvincing suggests that the rationalizing background story plays more than a rhetorical role in the deviant cases (cf. Heal 1999, 78; Boghossian 2011).⁸

Social externalism, then, is not the right explanation of holism. A second possibility is that deviant understanding is explained by a *causal theory of meaning* (Fodor 1987). On such a theory, for my use of some term *t* to mean some concept *c* is for my use of *t* to stand in the appropriate kind of causal relations with certain bits of the world: for example, my use of ‘sofa’ means *sofa* because it is causally connected in the right way with sofas. To explain the deviant cases, we have to say in addition that, if I use any term *t* that means some concept *c*, then I count as understanding *c*: for example, if my use of ‘sofa’ means *sofa*, then I understand sofa. So it’s possible to understand *sofa* while denying that sofas are for sitting on, as long as you have the right causal relations to sofas.

Moving to a causal theory of meaning would be a drastic reaction to the puzzle

⁸This is not to say that such bare cases are never possible – one such is Burge’s (1979, 77) example of someone who believes they have arthritis in their thigh. But here the overlap with orthodoxy is much greater than in the cases I’m considering.

posed by deviant understanding. For it would amount to abandoning not only inferentialism but also holism, and even the broad claim that understanding requires a setting. On the causal theory, it ought to be possible for a subject to have a single thought p as the entirety of their mental life, as long as the right causal relations hold. p might even be a necessary falsehood: there might be a subject whose entire mental life consisted of judging that green is red. This suggests that adopting the causal theory would be an overreaction. Nothing in the cases of deviant understanding puts pressure on the claim that understanding requires a setting. We should seek a more conservative explanation of these cases that doesn't require us to abandon what was plausible in our starting point.

The causal account also faces the same problem as the externalist one: it fails to explain the role of rationality in the cases. If causal relations were driving the cases, the background story would be beside the point. It would be sufficient to say: one day somebody believed that sofas are not for sitting on, while standing in the appropriate causal relation to sofas, and still understanding *sofa* in virtue of this relation. But, as before, this bare case is wholly unconvincing. So causal relations are not the right explanation either. Of course, this doesn't mean that causal relations have no role in understanding. It only means that, if they do have a role (as I believe they do), this role is not sufficient to explain the possibility of deviant understanding.

1.3.2 A better explanation

Both of the accounts I've considered stumbled on a distinctive feature of deviant understanding: the cases present a subject as being rational even in their irrationality. Why not make this feature central to our explanation? What underwrites understanding in these cases is precisely that the subject manifests rationality in the position that they take up. The background story rationalizes their mistaken belief by showing that it is supported by their other beliefs. This is why, in rejecting an obvious inference, they remain intelligible.⁹

The thinker who denies that sofas are for sitting on is intelligible because they

⁹As Schroeter & Schroeter write, 'anyone with whom we can profitably participate in critical debate involving an evaluative term – whatever his initial substantive assumptions are – shares the same meaning' (2009, 21). In more Fregean terms, a central theoretical role for thoughts is in tracing patterns of rational agreement and disagreement. As long as the deviant speaker shows sensitivity to rational relations, we should count ourselves as in rational disagreement with them, and this requires taking them to reject the very thought that we accept.

take this thought to be incompatible with a series of further thoughts which they accept: that sofas are religious objects, and that religious objects are not for sitting on. In other words, they reject a thought p because they accept some other thoughts q and r , which they rightly take to be in tension with p . We could analyze the rest of the background story in the same way. So the thinker's rejection of p manifests their ability to respond to the rational relations among their thoughts.

Exactly what ability does the thinker need? You might think that they only need to recognize particular rational relations – namely, those which form part of the ‘plausible background story’. In our example, they would only need to recognize the relations between p , q , and r – perhaps by having an aversion to believing p , q and r together. In fact, however, this suggestion reflects a residual inferentialism and fails to appreciate the depth of the holist challenge.

Positing an aversion to believing p , q and r together would account for the particular case I presented earlier, but it wouldn't account for the indefinitely many other cases we could imagine. In supposing that the deviant thinker understands *sofa* in virtue of the plausible background story, we suppose that they are sensitive to various further connections involving *sofa*: for example, if they rejected some crucial element of the background story about the propaganda campaign, they would come to accept that sofas are for sitting on – unless they replaced the propaganda campaign with a new background story – unless some crucial element of this new story proved incompatible with their other beliefs – and so on. There's an indefinite list of further connections to which a thinker who understands p is sensitive, an indefinite list of abilities to make further revisions in their beliefs as their other beliefs changed. In taking the subject to understand *this is a sofa*, we take them to have an open-ended sensitivity to the way other thoughts might bear on this one.¹⁰

Earlier, I treated the difference between inferentialism and holism as one of scope: the inferentialist holds that there is a set of rational relations involving p which you must accept to understand p , while the holist holds that to understand p , there must be a set of rational relations involving p which you accept. Now we can see that this understates the difference. There would be no deep challenge to inferentialism in allowing that there are, say, two sets of canonical inferences involving p , such that understanding p requires accepting at least one of these sets. (For example, there might be two sets of equivalent inference-rules for a connective.) The holist position

¹⁰Horvath (2020, 522) argues, relatedly, that understanding is not constituted by any particular first-order disposition but by a second-order disposition to have various first-order dispositions.

is, rather, that understanding p is an open-ended exercise requiring sensitivity to whichever other thoughts might prove to be rationally relevant.

The open-ended sensitivity involved in thinking means that no finite set of abilities to respond to particular rational relations will suffice for understanding. An ability to believe p if you believe q , for example, is entirely inert when it comes to other rational relations that may bear on p . What a thinker needs, instead, is a general ability to recognize and respond to rational relations involving p , such that their pattern of attitudes as a whole comes out as overall rational. To put it in a thesis:

General Rational Abilities A thinker must have general abilities to recognize and respond to relations of support and tension among thoughts.

To be clear, the requirement isn't only that thinkers have concepts applicable to cases of support or tension, nor even that they be able to recognize such cases reliably. They must also have some tendency to be moved by such relations to change their beliefs. If you recognize that p supports q , you must have some tendency to infer from p to q ; if you recognize that p is in tension with q , you must have some tendency not to judge p and q together.

1.4 The expressive role of logic

In the previous section, I suggested that thinking involves general abilities to recognize and respond to relations of tension and support among thoughts. Several philosophers have inferred from this that thinking requires metarepresentation, or second-order thinking about thinking. For example, Burge (1996, 100) argues that 'critical reasoning requires thinking about one's thoughts'. Spelling out Burge's argument, Rattan (2002, 150) writes:

I change my thoughts immediately when there is a tension in *my* thoughts, *when the perspective of thinking about thoughts and the first-order thoughts themselves are part of the same perspective*. So, it is because I normally know what my thoughts are, and that they are my thoughts, that reflection can allow us to take rational control of our thoughts. So special second-order knowledgeable thought seems to be essential to the very idea that we can exert rational control over our thoughts.

Ichikawa & Jarvis also argue that understanding involves topic-neutral inferential abilities to systematize our first-order inferential abilities; they take the systematizing

abilities to be second-order (2013, 90ff.).¹¹

It's not hard to see why one might take metarepresentation to be required. If we ask what the ability to recognize tension among thoughts consists in, a natural suggestion is that it consists in having the concept *tension*, and being able to use it in thoughts such as *p is in tension with q*. But to think *p is in tension with q*, you have to refer to – that is, metarepresent – the thoughts *p* and *q*. The same goes for *support*. So if the natural suggestion is right, then recognizing tension and support requires the expressive resources for metarepresentation – for second-order thinking about thoughts.

The problem is that this seems far too intellectualist. It seems like the ability to use that-clauses, quotation, or other devices of semantic ascent is much more sophisticated than the ability to think purely first-order thoughts (Price 1990, 231; Camp 2009, 286).

This creates a puzzle. On the one hand, thinking requires a general ability to revise one's position in response to rational relations among thoughts. On the other hand, thinking can't require a general ability to represent one's own thoughts. So there has to be a way of recognizing and responding appropriately to rational relations without metarepresentation. In 1.4.1, I'll argue that logical concepts – in particular, negation and the conditional – play this role. Following this, in 1.4.2 I sum up the argument and comment on where I diverge from Brandom's views.

1.4.1 The conditional and negation

In this section, I argue that having logical concepts – specifically, the conditional and negation – provides a way of being responsive to the rational relations among thoughts without needing to metarepresent them.

Expressive Role of Logic The conditional and negation allow their possessors to have general rational abilities without metarepresentation.

The basic idea is simple. Instead of having a meta-level concept like *support*, which forms a thought from two names of thoughts, you have an operator \star which forms a thought from two thoughts, where $p\star q$ is true if *p* supports *q*. Given this fact, being able to think $p\star q$ is a way of being able to think about the support-relation between

¹¹Although in other places their description of systematization sounds less metarepresentational (2013, 228). For discussion of this tension, see the review by Rattan (2014).

p and q without metarepresentation. Of course, \star just is the conditional. A similar story goes for tension and negation.

This basic idea isn't new. It can be found in the Grammatical Propaedeutic to Anderson & Belnap (1975), which argues for a systematic equivalence between operators within a language and metalinguistic predicates which express facts about the language. Formal presentations of logical connectives as reflecting metalinguistic properties in the object-language can be found in Došen (1989) and Sambin, Battilotti & Faggian (2000). Perhaps the most detailed philosophical development of the idea is by Brandom (1994), who also singles out negation and the conditional as exemplifying this role.

This idea about the expressive role of logic doesn't dictate exactly which logic is correct. For purposes of clearer exposition, however, I'm going to assume that there's a single set of correct rules, and will ignore the existence of controversy about which rules those are. I won't attempt to fully specify which rules those are, but for each concept I will suggest a minimal rule that the concept must satisfy. In the next chapter, I consider how my account would have to change to accommodate controversy about which rules are correct and the pluralist view that multiple logics are correct.

I'll discuss the conditional first and then negation.

The conditional. Possessing the conditional constitutes our general ability to recognize and respond to relations of support among thoughts. If *the oven is too hot* supports *the pie will burn*, then we can express this relation by thinking, *if the oven is too hot, the pie will burn*. The latter is an object-level thought; thinking it doesn't require the ability to metarepresent either of its constituents. More generally, if p supports q , we should have *if p , then q* .

In positing such a close link between the conditional and the consequence-relation, I'm following Došen (1989, 366) and Priest (2006, 83). But the consequence-relation and the conditional are not the same in every respect. They're of different logical types – the former is a metalanguage predicate, the latter an object-language operator – which gives them different uses. Given this difference, Shapiro (2011) argues that the consequence-predicate can be used to formulate generalizations that are not expressible in the object-language, in much the same way the truth-predicate can be used for formulating generalizations that are not expressible in the object-language. For present purposes, the more important difference is that the conditional can be used by a subject who lacks the resources for metarepresenting their thoughts.

A defining feature of relations of support (or at least the strong relations I've focused on) is that if p supports q , then from p , you can infer q . If a conditional serves to express these relations, then it works as an 'inference-ticket' licensing its possessor to move from p to q (Ryle 1949, 117). In other words, the conditional has to satisfy something like Modus Ponens:

Modus Ponens From p and *if p , then q* , infer q .

Priest (2006, 83) even writes that Modus Ponens is 'analytically part of what implication is'.

I say 'something like' Modus Ponens because McGee (1985) has proposed counterexamples to Modus Ponens; I want to allow for the possibility that only a restricted version of the rule is required. But even the reasoning which McGee uses to make his case involves instances of Modus Ponens: for example, that if Modus Ponens has a counterexample, it should be modified, and Modus Ponens does have a counterexample, so— I don't take this to show that McGee's argument is incoherent. Rather, I take it to show how unavoidable Modus Ponens, or something like it, is in making a case for anything at all (Russell 1912, ch. VII; Evinine 2001, 348-9; Hale 2002).

Although I've stated Modus Ponens as an imperative, it doesn't mean that if you judge some p which supports q , judging q is all things considered the best thing to do. You might also believe some t which is in tension with q and more credible than p . This doesn't undermine the claim that support-relations are expressed by a conditional which satisfies Modus Ponens. As I'll explain in more detail in Chapter 3, following a rule doesn't mean always conforming to it; it means tending to find the conclusions of instances of the rule compelling on the basis of instances of the premises. Following the rule in this sense is consistent with judging p and not judging q in some cases.

Note, too, that not every conditional has to play this expressive role. It may be that many conditional constructions in natural languages play other roles (see Briggs (2019) for general discussion of conditionals). Moreover, there's no guarantee that the conditional which does play this expressive role is a material conditional, definable as *not (p and not- q)*. Rather, the logical properties of the conditional will reflect the logical properties of the support-relation. So, for example, if Brandom (2018, 78) is right that support is not monotonic – if p might support q even though p and r does not – then the conditional might not allow strengthening the antecedent:

we could have *if p, then q* but not *if p and r, then q*.¹² If, by contrast, support is conclusive and therefore monotonic, strengthening the antecedent should hold (Ichikawa & Jenkins 2013, 25).

Negation. Possessing a concept of negation constitutes our ability to recognize and respond to tensions between thoughts. The role of negation is brought out vividly in a dialogue imagined by Huw Price. He imagines a community of Ideological Positivists, ‘fanatical disciples of Norman Vincent Peale’, who ‘have tried to reconstruct their language so as to make negative thinking impossible’ (1990, 223). The Ideological Positivists have, in particular, purged any term for negation from their language. Now Price imagines two Positivists discussing where to find their friend Fred:

Me: ‘Fred is in the kitchen.’ (Sets off for kitchen.)
You: ‘Wait! Fred is in the garden.’
Me: ‘I see. But he is in the kitchen, so I’ll go there.’ (Sets off.)
You: ‘You lack understanding. The kitchen is Fred-free.’
Me: ‘Is it really? But Fred’s in it, and that’s the important thing.’
 (Leaves for kitchen.)

The point of the dialogue, I take it, is that the characters – or at least ‘Me’ – lack even an implicit understanding of tension between claims.¹³ ‘Me’ seems to accept that Fred is in the garden, and is unable to appreciate that this is in tension with Fred’s being in the kitchen. As Price comments (1990, 224):

Your problem is to get me to appreciate that your claims are incompatible with mine. Even in such a trivial case, we can see that it would be useful to have a device whose function was precisely to indicate that an incompatible claim was being made: precisely to deny an assertion or suggestion by somebody else. It seems that this is what negation gives us.

If the characters in the dialogue understood negation, they could say things like

¹²Brandom makes the stronger claim that, since the conditional serves to express a prelogical relation of material support, it should be conservative over the prelogical support-relation. So if prelogical support is non-monotonic, the conditional has to be non-monotonic. In my own account, the link between the conditional and support is different. The conditional helps to usher in a new kind of representational activity; there’s no reason in principle why it should be conservative over a prelogical support-relation rather than modifying it (see also 3.5.2). But its role is still to express support. So if the *post*logical support-relation is non-monotonic, then the conditional should be too.

¹³It might be argued that ‘You’ displays such an implicit grasp in the use of language such as ‘lack’ and ‘Fred-free’. This is fine with me: thinkers need to understand negation, but this doesn’t require that they have a single word that plays the same role as ‘not’ in English.

‘Fred is not in the kitchen’; this claim is such that, if you understand it, then you take it to be in tension with ‘Fred is in the kitchen’.

How exactly does negation express tension between thoughts? The essential feature is that *not-p* and *p* are in tension, or mutually exclusive (Ervine 2001, 354). But of course many propositions are in tension with *p*. More precisely, then, the essential feature is that *not-p* is the weakest claim in tension with *p*: if *q* is in tension with *p* then *q* entails *not-p* (Peacocke 1987, 163; Brandom 1994, 115).¹⁴

Now, the defining feature of tension between thoughts is that if *p* and *q* are in tension, you shouldn’t judge *p* and *q*. So if *not-p* is the weakest claim in tension with *p*, then, as a minimum requirement, it must satisfy something like the Law of Non-Contradiction in the following form:

Law of Non-Contradiction Do not judge *p* and *not-p*.

As Hale suggests, ‘acceptance of the principle of non-contradiction may be integral to operation with the idea that recalcitrance obliges us to make some revision in our overall corpus of accepted statements’ (1999, 51 n. 24).

It’s true that dialetheists have argued against Non-Contradiction (Priest 2005; 2006), generally on the basis that rejecting Non-Contradiction makes room for a more satisfactory solution to the semantic paradoxes. But the very idea that the paradoxes could motivate such a rejection presupposes a notion of rational tensions among our commitments which demand resolution (Armour-Garb 2004). For example, the reasoning might be that if the semantic paradoxes give us reason to accept *p* and *not-p*, and the Law of Non-Contradiction tells us not to accept *p* and *not-p*, then we ought to give up one or the other of these commitments. This is an instance of something like the Law of Non-Contradiction, even if we only want to allow a restricted version of the rule. Again, this doesn’t show that these arguments for dialetheism are incoherent, but it does show how unavoidable something like Non-Contradiction is in arguing that there is a tension among our beliefs which requires us to make a revision.

¹⁴This doesn’t dictate precisely which logic of negation to adopt. For an analogue of this idea in relevant logic see Restall (1999, 61). Peacocke (1993, 177) argues that if *not-p* is the weakest claim in tension with *p*, then negation is classical. But Wright (1993) shows that Peacocke’s argument presupposes either Double Negation Elimination or a classical conception of the domain of thoughts. For further discussion, see Humberstone (2011, 1170-1171).

1.4.2 Response to Brandom

That concludes my argument for **Logical Concepts**. There were a lot of moving parts, so it might be helpful to recap before we move on. Here's how the argument went:

1. (**Holism**) Understanding a thought p requires sensitivity to some of its rational relations to other thoughts.
2. The best explanation for Holism is that thinkers must have general abilities to respond to support and tension among their thoughts. So
3. (**General Rational Abilities**) A thinker must have general abilities to respond to support and tension among their thoughts.
4. But thinkers don't need metarepresentation.
5. (**Expressive Role of Logic**) The conditional and negation allow their possessors to have general rational abilities without metarepresentation.
6. There's no way other than possessing the conditional and negation to have general rational abilities without metarepresentation. So
7. (**Logical Concepts**) A thinker must possess concepts of the conditional and negation.

Since, as I've argued, the conditional and negation have certain minimal rules, I take it to follow from Logical Concepts that thinkers have to be sensitive to logical rules. I discuss the kind of sensitivity involved in Chapter 3.

The only premise I haven't defended is **6**. I can't think of a clear counterexample to it. Of course, a thinker might be able to use other concepts from which negation and the conditional can be defined: if they were using classical logic, they might possess (say) conjunction and negation, or a Sheffer stroke. **6** should be read to allow for such cases.¹⁵ But it's important to notice that which equivalences are available will depend on which logic is correct: for example, in the Logic of Paradox (2.4; Priest 2006), if we define *if p , then q* as *not (p and not- q)*, Modus Ponens fails. So in this logic, conjunction and negation don't provide an alternative to the conditional.

I'm now in a position to respond to Brandom's claim that there is 'nothing incoherent about a language or stage in the development of a language in which the

¹⁵Not *every* functionally complete set is allowed. Consider Leśniewski's Protothetic, which uses only the biconditional and propositional quantification (Tarski 1923). Protothetic is functionally complete, but following its rules requires understanding propositional quantification. I'm inclined to think this requires metarepresentation, so it's not an alternative to the conditional and negation.

only vocabulary in play is nonlogical' (1994, 383). In terms of the above argument, Brandom accepts **1**, **4**, **5** and, I think, **6**. Because he rejects **7**, he must reject **2** and **3**. He concedes that logic 'make[s] it possible to criticize, control, and improve our concepts' (1994, 384), but he denies that this is required for thinking. His idea is that a prelogical thinker can be sensitive to rational relations between particular contents, but not to rational relations in general. Such a 'thinker' would lack conceptual freedom.

To illustrate this point, let's consider the *affirmativist* – someone whose language lacks negation, but contains for each primitive predicate F a primitive predicate \bar{F} for its complement (Kripke 2015). For each F and \bar{F} , the affirmativist follows a rule not to predicate both of a single object. Let's also suppose (modifying Kripke's example) that the affirmativist lacks other logical connectives, such as the conditional, but that they follow rules specifying containment between primitive predicates: if F contains G , then if you predicate F of an object you must also predicate G of that object. The affirmativist is sensitive to support and tension between particular thoughts, but they have no topic-neutral way of expressing these rational relations.

Every rational relation we recognize can be hardwired in to the affirmativist's language, so the set of relations they recognize can be the same as ours. But they have no way of proposing to add or drop a rational relation. Supposing there isn't already a rule that being F supports being H , the affirmativist cannot propose *If Fa, then Ha*. Supposing there's a rule that being F is in tension with being I , the affirmativist cannot propose *Not: If Fa, then not Ia*. But proposing to add or drop rational relations is precisely what we do in exercising conceptual freedom; this is just what the deviant speakers discussed earlier were doing. So the affirmativist lacks conceptual freedom. Since Brandom is committed to treating the affirmativist as a thinker, he has to deny that conceptual freedom is required for thinking.

In a response to Brandom, John McDowell writes that 'self-consciousness requires the capacity to make the goodness of materially good inferences explicit, and hence command of logical vocabulary; and it is unintelligible that something without (semantic) self-consciousness could explicitly undertake commitments' (1997b, 162). In later comments, McDowell elaborates, pointing to the idea that grasp of meaning requires 'being responsive to reasons' (2005, 135):

Surely the responsiveness to reasons that figures in this connection should be responsiveness to reasons *as such*. Can that really be in place in the absence of the capacity to raise questions about whether what one finds

oneself inclined to be swayed by, in forming a belief or deciding to act, really constitutes a reason for the belief or action one is contemplating?

McDowell is right to take issue with Brandom's claim, but the way he does so is too intellectualist. He seems to be suggesting that thinking needs logic because it needs metarepresentation, and metarepresentation uses logic. While I agree that thinking needs logic, my reason is nearly the opposite of McDowell's: thinking needs logic because it *doesn't* need metarepresentation.

The fact that thinking needs logic doesn't mean that every time a thinker judges q on the basis of p , they use the conditional *if p , then q* , or that any time a thinker avoids judging p and r because they are incompatible, they first infer *not- p* from r . I agree with Brandom that material inferences, which don't use logic, need not be enthymematic. But it doesn't follow from the fact that material inferences don't *use* logic that logic makes no difference to their nature (McDowell 2005, 134). The assumption that it makes no difference reflects what Koreň (2018, 201) calls a 'layer-cake' picture of language, in which logic is added on top of an existing practice without transforming it in any way.¹⁶

But logic does make a difference. The conceptual practice of a subject without logic would involve sensitivity to particular links between thoughts. But nothing in their practice would allow them to reject such a link while maintaining understanding. Nor, therefore, would they be able to understand someone else who rejected the link. As a result, they would relate to the rules that they followed as a rigid structure imposed externally, not as a set of apparently rational connections which can be modified as rationality requires. We need not deny that such a speaker would be able to say intelligible things. But the rigidity of their practice means that they would not manifest the rationality that (if my line of thought is correct) is essential to thinking.

1.5 Quinean reprise

At this point, I think it might be useful to take a different approach to the material we've just covered. The holism about understanding which we find in Burge and Williamson is a philosophical descendant of Quine's holism about confirmation. I'll

¹⁶Boyle (2016) develops a related critique of 'additive' theories of rationality, although his focus is on the way rationality transforms perception and desire.

show that the role of logical rules, which I've been trying to excavate, is much more visible in the Quinean picture.

While there's a lot of disagreement about how to understand Quine's arguments in 'Two Dogmas of Empiricism', we can set that aside to focus on the positive proposal near the end of the paper: a holistic epistemology in which no statement is immune from revision. Quine writes (1951, s. VI):

The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Re-evaluation of some statements entails re-evaluation of others, because of their logical interconnections – the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having re-evaluated one statement we must re-evaluate some others, whether they be statements logically connected with the first or whether they be the statements of logical connections themselves. ...

Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision.

Supposing that an analytic statement must be immune to revision, it follows from Quine's picture that there are no analytic statements. Some statements, like those of logic, are central while others are at the periphery; in principle, however, any statement can be revised or held fixed in the face of any experience.

This picture is the ancestor of that of Burge and Williamson. A statement like 'Sofas are not for sitting on' can be held true, or a statement like 'Every vixen is a vixen' can be rejected, in the face of a great deal of recalcitrant experience 'if we make drastic enough adjustments elsewhere in the system'. Indeed, Williamson's examples nicely show how revising a belief deep within the web – e.g. about the semantics of 'every' – can ramify into apparently unrelated domains. Quine, Burge and Williamson all allow us a freedom in our use of concepts that results from the

‘intricacy’ (Quine 1974, 38) of the web of belief. Given this connection, if logic turns out to have an essential role in Quine’s account of the web of belief, we have some reason to suspect that it will turn out to be similarly essential in contemporary holism about understanding.

The role of logic in Quine’s web is more subtle than is often appreciated. Logic appears in two guises: first as ‘logical interconnections’ among beliefs, such that revising one belief can mandate revising others, and second as ‘statements of logical connections’ (‘if the oven is too hot, the pie will burn’) and ‘logical laws’ (‘From p and *if p, then q*, infer q ’). Statements of logical connections and logical laws, in Quine’s view, are simply further nodes of the web, but logical interconnections themselves cannot be. This subtlety, however, is often overlooked. Quine is taken to assign to logic only the second role, to hold that logic occurs in the web only as nodes, not as connections among nodes. But if logic only occurs as nodes in the web, then it isn’t a web at all, but merely an array of disconnected points. I’ll begin my discussion by rehearsing the problem for taking logic to consist only in ‘further statements of the system’ and then comment on the other role that logic must play.

The claim that logical laws are ‘further statements of the system’ faces an initial difficulty: ‘logical laws’ is most naturally heard as referring to rules of inference, not propositions (Dummett 1973, 596; Priest 1979, 291). But a rule of inference, like Modus Ponens, is not the sort of thing that can be believed or disbelieved; rather, it is the sort of thing we can follow or fail to follow. So it can’t be a node in the web of belief. Still, for any rule of inference of the form ‘from Γ , infer q ’, we can write down a closely related statement: the corresponding conditional, ‘if Γ , then q ’ (Priest 1979, 292). So we can take the claim to be that the role of logical laws in the web of belief is exhausted by the occurrence of their corresponding conditionals as nodes within the web.¹⁷ I’ll refer to this as *condensing* a rule into a truth.

The common reading of Quine holds that logical laws occur in the web only in condensed form. The problem with this picture is that it makes it impossible for any belief to support any other. Suppose, for example, that the web contains the beliefs p , *if p then q* and q . The first two beliefs should count as supporting the third – but how does this happen? The only answer available is that the web contains

¹⁷Some have argued that all rules of inference are of conditional form (Finn 2019b), but this view isn’t mandatory. If we also want to allow prohibitory rules, like ‘Do not believe p and *not-p*’, we should extend the notion of corresponding conditional to include statements like ‘not p and *not-p*’. If we want to allow rules of inference to have multiple conclusions (Restall 2005), we should extend it to include statements like ‘if p , then q or r ’. See also 2.4.

a corresponding conditional for Modus Ponens, as an axiom-schema or statement universally quantifying over thoughts, or for a particular instance of Modus Ponens. But if p and *if p , then q* are not enough on their own to support q , then adding such a conditional – simply a further if-then statement – will not improve the situation. As Priest and Dummett both observe, the assumption that an added conditional is what does the supporting leads to the regress in Carroll (1895).

The same is true of tensions between beliefs, whose existence Quine needs in order to hold that a ‘recalcitrant’ experience forces us to make some adjustment or other in our beliefs. Suppose that the web contains r and *not- r* . These two beliefs should count as being in tension – but how does this happen? The only answer available is that the web also contains something like *not r and not- r* . But if r and *not- r* are not already in tension, adding this further belief will not improve the situation. As BonJour writes (1998, 94),

the basis for any supposed incompatibility within any set of sentences ... can apparently only be some further sentence in the system that says explicitly that the acceptance of such a set is objectionable and hence that the system of beliefs must be revised. But if we now consider the set of sentences that includes that one, the same situation repeats itself...

In short, if statements already stand in relations of support and tension, then corresponding conditionals are not required for underwriting those relations; if they don’t already stand in such relations, then corresponding conditionals won’t make a difference, and all we have is a ‘featureless collection of sentences standing in no special relations’ (Dummett 1973, 597).

It follows that statements must already stand in relations of support and tension. In other words, corresponding conditionals can’t exhaust the rational role of logical laws in the web of belief. Logical laws also have to provide the structure of the web. Let’s consider how this works.

Quine’s claim is not merely that any belief can be held fixed or rejected in the face of experience, but that any belief can be rationally held fixed or rejected in the face of experience (Chalmers 2011a, 389). The fact that an analytic truth can be rejected irrationally would pose no threat to defenders of analyticity. Quine’s web shows how, given enough compensatory adjustments, any p can be held fixed or rejected while the subject remains rational – or at least rational enough to count as understanding p . But this only makes sense if there’s a fact of the matter, to which the subject is responding, about which combinations of other beliefs should

be accepted or rejected if p is accepted or rejected (Wright 1986a; Field 1998, 13).

There may be room for debate about whether p is true, and about whether p is in tension with, or supports, some particular q . But if logical laws themselves are up for grabs – if it’s up for grabs whether p and *if p , then q* support q – then there’s no sense in which the subject ought to accept q if they accept p and *if p , then q* and so no sense in which it would be rational to make such an adjustment; it would be equally rational to accept the premises and reject the purported support they provide to the conclusion. In this case we might say with Wittgenstein that ‘whatever is going to seem right to me is right. And that only means that here we can’t talk about “right”’ (*Philosophical Investigations* §258). The viability of the web conception, then, requires that there be objective logical rules, which the subject is able to recognize and respond to appropriately.

You might think that it would be enough for the subject to recognize and respond to particular rational relations – for example, those between p , q and r – without following any general rules. But this fails to allow for the open-ended character of the phenomenon. To say that the subject is rational in accepting p is not merely to say that they responded appropriately to the relations between p , q and r , but to say that there is an indefinite list of further changes that they are capable of making, depending on which other beliefs turn out to be relevant. So the thinker needs a general ability to respond to relations of support and tension such that their pattern of attitudes as a whole comes out rational. And this just is an ability to follow logical rules.

This, in turn, casts doubt on Quine’s claim that logical rules are all up for rejection. There’s room for debate about exactly which rules are the right ones – for example, whether *not-not- p* supports p . But it’s not the case that any set of rules will do. Any revision, if it is to be rational, has to rely on an intelligible conception of support and tension among beliefs; if we change our logic so much that we no longer have a conception of support and tension, then no set of beliefs is rational at all. There’s no substitute for support and tension, no alternative relations that can structure a web of belief. Nor, therefore, is there a substitute for the logical rules which govern support and tension. The fixity of these relations makes possible the freedom we see elsewhere in the system. As Wittgenstein said (*On Certainty* §343): ‘If I want the door to turn, the hinges must stay put.’

1.6 Conclusion

I've argued that logic is necessary for thinking in the sense that understanding any thought requires possessing logical concepts, which means being sensitive to logical rules. This argument, if successful, establishes not only that a subject with no sensitivity to logical rules would be unintelligible to us, or that we would have no reason to attribute thinking to them, but that, objectively speaking, their representational activity would not be thinking at all. Similarly, it shows not only that 'we have no choice but to project our own logic on to the beliefs of another' (Davidson 2004, 157) but that both we and the other must share the correct logic.

But the strength of this conclusion should not be overstated. It shows that all thinking is logical, but not that all representational activity is logical: we already know that there are kinds of representational activity, like sensing and perhaps perceiving, which are not necessarily logical. The kinds of non-logical representational activity which we know of might seem to be less objective, or at least less sophisticated, than thinking, but it's an open question whether there could be creatures whose representational activity is as objective or as sophisticated as thinking without involving logical rules. If such creatures were possible, then the fact that logic is necessary for thinking would turn out to be more parochial than it sounds.¹⁸

¹⁸This possibility arises in Kant's account of thinking. In the *CPR*, Kant maintains that the logical categories are the universal structure of thinking; although other thinkers might have different forms of sensibility (B150), they could not have different (or no) categories. Later on, in the *Critique of the Power of Judgment*, Kant observes that the claim that the categories are universal is more parochial than it sounds: all we can really say is that 'given the nature of our (human) cognitive faculty or even the concept that *we can form* of the capacity of a finite rational being in general, we cannot and must not conceive otherwise, but without asserting that the basis for such a judgment lies in the object' (§76; AK 5:401). Similarly, the argument of this chapter relied on our conception of thinking for its conclusions about logic; I did not attempt to show that the world itself grounds these conclusions because it can only be thought about logically. As a result, the possibility of a similarly sophisticated but non-logical representational activity is left open. As Kant writes, we can conceive this possibility only 'negatively, namely merely as not discursive' (§77; AK 5:406), observing only that nothing said so far rules it out.

2 Disagreement and pluralism

2.1 Introduction

In Chapter 1, I assumed that there's a single correct set of logical rules – in short, the One True Logic – and ignored the existence of controversy about which rules these are. These commitments allowed me to focus the argument, but they are highly disputable. First, even if there is one true logic, there is longstanding disagreement about which logic it is: candidates include classical logic, intuitionistic logic, relevant logics, dialethic logics, and other three-valued and four-valued logics. Second, logical pluralists have argued that there is more than one correct logic (Beall & Restall 2005; Russell 2008; Shapiro 2014).¹ In this chapter, I address the ways that my account needs to be modified if we relax the initial commitments to allow for controversy about which logic is correct and for logical pluralism.

Let me briefly restate the results of Chapter 1. I argued that thinkers must possess concepts of the conditional and negation. More fully, a subject's representational activity counts as thinking only if that activity manifests sensitivity to the rules that govern the conditional and negation. The upshot, which I'll discuss in detail in Chapter 4, is that there are limits on illogical thinking: if you don't manifest sensitivity to the logical rules, you fail to think at all. What survives of these claims if we relax our commitment to a single and uncontroversial set of logical rules?

To pin down the issues, we have to draw a connection between a deductive system such as classical or intuitionistic logic and the rules that a thinker must follow. You might think that these have, at best, a distant relation to each other: deductive systems are precise, while the rules we think with are vague. Indeed, I emphasized

¹In some ways, everybody should be a pluralist: everybody should accept that there are different abstract structures with associated model or proof theories – what Priest calls 'pure logics' (2005, 195). Moreover, everybody should accept that different pure logics might be best for different applications: one for computer programming, another for constructing mathematical proofs. These tame pluralisms pose no threat to anyone. For a pluralism in which logic can vary with the kind of truth-bearer, see Russell (2008), and for some other varieties see Shapiro (2014, ch. 2). The challenging pluralism at issue here holds that, vis-a-vis some important role a logic could play, more than one logic is correct for that role. For example, Beall & Restall (2005) argue that more than one logic specifies an 'admissible' consequence relation – where a consequence relation is admissible iff it is necessary, formal, normative and instantiates the Generalized Tarski Thesis ('An argument is valid_x if and only if in every case_x in which the premises are true, so is the conclusion'). My focus here is on the role of constituting thinking. The pluralism I'm considering is pluralism about *that*.

in 0.3.1 that there was no upfront guarantee that logic, in my sense, has anything to do with deductive systems. I didn't want to assume, for example, that the rules we think with are identical to those of classical logic. But I now want to suggest a connection between deductive systems and the rules we think with (cf. Blake-Turner & Russell 2018, s. 3). To make this connection, I'll need to say a bit more about deductive systems and about how thinkers follow logical rules.

I take a *deductive system* to be composed of three things:

- An artificial object-language;
- A set of *entailments* of the form $\Gamma \vdash Q$, which say that some (possibly empty) set of object-language sentences Γ entails some sentence Q ; and
- A *superstructure* which generates the entailments – typically, a model theory or proof theory (for further discussion, see Russell (2019, 552)).

Of course, in practice we don't always specify a superstructure, and I won't always do so here; in principle, however, the choice of superstructure can make a difference where there are failures of soundness or completeness. I will say that $\Gamma \vdash Q$ is valid in a deductive system if and only if it belongs to its set of entailments. To be clear, this means I'm using the single turnstile in a broader way than is usual: it expresses entailment, whether generated by model-theory or proof-theory. On a model-theoretic approach, an entailment holds iff Q is true (or: has a designated value) on condition that Γ are true (or: have designated values). On a proof-theoretic approach, an entailment holds iff there is a proof of Q from Γ , either in a single step or in a series of steps.

We follow logical rules by having logical capacities – capacities to make certain transitions between thoughts. As I'll explain in detail in Chapter 3, someone who possesses a logical capacity finds transitions which obviously fall within the relevant rule 'primitively compelling'. For example, someone who has the capacity to draw Modus Ponens inferences will find it primitively compelling to go from p and *if p , then q* to q . Let us say that a form of inference is 'licensed' by a thinker's logical capacities if the form of inference is either primitively compelling or 'derivatively' compelling – that is, it can be built up out of primitively compelling steps.

To approach the connection between deductive systems and the rules thinkers follow, let's consider the following form of inference:

From *not-not- p* , infer p .

It seems like an open question whether this form of inference is licensed by the logical capacities which thinkers must possess. It seems doubtful that it is primitively compelling. For it seems possible that someone could judge *not-not-p*, fully understanding it, and yet not take themselves to have a reason to judge *p* even when the question is raised. It does, however, seem like a live possibility that the rule is derivatively compelling.

Now, notice that while classical and intuitionistic logic both specify rules for negation, these rules differ: the following is a valid form of argument in classical logic and not in intuitionistic logic.

Double Negation Elimination $\neg\neg P \vdash P$

The fact that DNE is a valid form of argument in classical logic and not valid in intuitionistic logic points to the connection between deductive systems and the rules we think with. If thinkers are licensed to infer from *not-not-p* to *p*, then, to this extent, classical logic is correct and intuitionistic logic is incorrect. More generally, then, a deductive system is correct if and only if the entailments in the system specify the inferences licensed by thinkers' logical capacities.²

Let me enter a few caveats. First, I'm not claiming that this is what mathematical logicians have in mind or that this is the only right way to see deductive systems.³ Second, the claim is that the class of transitions licensed by (say) classical logic corresponds to the class of transitions licensed by thinkers' capacities – not that each rule in a particular presentation of classical logic corresponds to a particular capacity. So the correctness of a deductive system doesn't require that the rules in the deductive system have psychological reality. Third, then, deductive systems that validate the same entailments will stand or fall together. If classical logic is correct, then a system with only the Sheffer stroke is as correct as a system with the conditional and negation.

We've now established a connection between deductive systems and the rules we think with. With this connection in place, we don't need to be fussy about the distinction between logics and deductive systems. Going forward, I'll talk about

²Of course, in classical logic negation and the conditional are an expressively complete set (Enderton 1972, §1.5), so accepting classical rules for those connectives amounts to accepting classical logic. This is not so in, for example, intuitionistic logic. So in principle someone might advocate accepting the rules that some system contains for negation and the conditional without those for, say, disjunction. For readability, however, I'm going to assume that the candidates for the correct logic are relatively familiar deductive systems and not fragments of them.

³On other cognitive attitudes one might have to a deductive system, see Soysal (2018).

whether there's a single correct 'logic' or many, or about someone advocating or using a 'logic': all of these locutions can be cashed out in terms of deductive systems which are said to capture the rules we think with.

Having clarified this point, we can restate the commitments we started with: first, the metaphysical assumption of a single correct logic for thinking, and second, the epistemic assumption that this logic is universally accepted as correct by thinkers.

Monism There is only one correct logic for thinking.

Agreement The correct logic is universally accepted as correct.

Relaxing these assumptions allows for the following:

Pluralism There is more than one correct logic for thinking.

Disagreement The (or a) correct logic is not universally accepted as correct.

For now, I'm going to ignore the possibility that *no* deductive system correctly specifies the logical rules that thinkers have to follow. This possibility would obtain if there were no such rules – a variety of 'logical nihilism' (Russell 2018a) – but it would also obtain if, while there were such rules, no deductive system was adequate to specify them (Cotnoir 2018). Perhaps, as Strawson said, ordinary language 'has no exact logic' (1950, 344). I'm ignoring this view because the more difficult problems for my view involve Pluralism and Disagreement.

We can organize the space of options in stages, as follows:

1. Monism and Agreement
2. Monism and Disagreement
3. Pluralism and Disagreement

Possibility 1 is what I've been working with so far in the dissertation. In this chapter I'll discuss what changes I need to make to accommodate Possibilities 2 and 3. There's also the theoretical possibility that many systems are correct, but only one is universally accepted. In this case, the pluralism would be possible but not actual. I don't think any new issues arise in this case.

2.2 Monism and Disagreement

At this stage we retain the assumption that there is a single correct logic for thinking, but we drop the assumption that this logic is universally accepted as correct. Rather, there is disagreement about which logic is the right one. This combination of views appears most prominently in the work of defenders of non-classical logics who argue on principled grounds that their logic is the right one. Dummett, for example, argued in *The Logical Basis of Metaphysics* (1991) that general constraints on the theory of meaning show that intuitionistic logic is correct. Similarly, Priest has argued that a dialetheist logic is correct on the basis that it allows a coherent treatment of various paradoxes; he denies (2005, ch. 12) that it is merely one of many correct logics. Defenses of classical logic against such challenges (Rumfitt 2015) also make the monistic assumption.

A similar position sometimes – but not always – underlies discussions of whether logic is rationally revisable. Haack (1996, 26) distinguishes between ‘absolutist’ and ‘pragmatist’ approaches to revision: the absolutist assumes that there is a unique correct logic, but argues that we have been wrong about which logic it is, while the pragmatist holds that we should choose the logic which displays the best combination of theoretical virtues. A Haackian absolutist is a monist who allows for disagreement about which logic is correct.⁴

On a view where logic has the same status as other sciences (Hjortland 2017), disagreement about the right logic can be as intelligible and rational as any other scientific disagreement. By contrast, you might worry that, if there is a single correct logic which is constitutive of thinking, then anyone who disagrees with that logic must be failing to think. You might also worry that, even if someone who disagrees with the right logic is still thinking, the rationality of that disagreement is cast into doubt.

In response, I want to begin by distinguishing two kinds of logical disagreement. Type A disagreement consists in having conflicting beliefs about which logic is correct. Type B disagreement consists in actually using conflicting logics. The distinction here is between what Peirce called *logica docens* – beliefs about logic – and *logica utens* – logic in use (Pietarinen 2005). We will see that on the constitutive view, advocating an incorrect logic does not lead to failure to think, nor to automatic

⁴A Haackian pragmatist could hold a similar position, if they thought that theoretical virtues were a reliable guide to the correct logic, or they could be a pluralist.

irrationality. In this sense, constitutivism is consistent with disagreement about logic. Actually using an incorrect logic may lead to failure to think, but I'll suggest that our discomfort with this conclusion is better targeted at monism than at constitutivism.

2.2.1 Type A disagreement: conflicting beliefs about logic

Type A disagreement occurs when the two parties have conflicting beliefs about which logic is correct: for example, R asserts that classical logic is correct and D asserts that intuitionistic logic is correct. There are two worries you might have about how this kind of disagreement interacts with my view.

The first worry is that someone who advocates intuitionistic logic must be failing to think. This worry can, I think, be assuaged fairly quickly. It's possible to advocate intuitionistic logic while following the rules of classical logic (or, conversely, to advocate classical logic using only intuitionistically valid reasoning). Indeed, this fact underlies the argumentative strategy of Rumfitt (2015), who argues for classical logic against intuitionism by providing a semantic theory that validates classical logic regardless of whether the metalogic is classical or intuitionistic.⁵ So the fact that someone is advocating an incorrect logic doesn't mean that they're failing to think. (If they're *using* an incorrect logic, that is a different matter, which I consider in the next section.)

The second worry is that someone who advocates an incorrect logic must be irrational. Field (2009, 251) suggests a similar objection, arguing that the claim that logical laws are laws of thought

seems problematic, if rational change of logic is possible: can it really be that in a debate over logic, the party who advocates the incorrect logic is automatically irrational, however compelling her case may be and however poor the currently available arguments on the other side?

But Field's argument is not entirely clear. It's trivially true that someone who advocates an incorrect logic is *wrong* – but why would they be 'automatically irrational'? One possibility is that Field assumes the advocate of the incorrect logic must be using the incorrect logic, and so reasoning incorrectly; but as I pointed out above, there's no reason why someone who advocates a logic has to use that logic in their advocacy.

⁵Rumfitt is following the strategy of Dummett (1991), who argues for intuitionism. Williamson (2017) argues instead that the metalogic should be the same as the logic, but my disagreement with Williamson is upstream of this particular issue.

Another possibility is that Field thinks logics are *immodest* in roughly the sense of Lewis (1971): good classical reasoning will inevitably favour classical logic, good intuitionistic reasoning will inevitably favour intuitionistic logic, and so on. The argument would then be that since (say) classical logic is constitutive of thinking, the advocate of intuitionistic logic must be following classical logic; but good classical reasoning favours classical logic; so the advocate of intuitionistic logic must be reasoning badly. This is an interesting argument, but – aside from doubts, suggested by the earlier examples of Dummett and Rumfitt, about whether logics are immodest in this sense – I don't think it is what Field has in mind. Field's imagined advocate is not reasoning badly, but well; they are making a compelling case against a not-so-compelling one.

Rather, I think Field's argument is something like this. If there is a set of logical rules which all thinkers follow, then a thinker who is wrong about what those rules are is misapprehending the very rules that they are following. Supposing that the right logic is classical, for example, Priest, Dummett and other non-classical logicians are wrong about the very rules which they, as thinkers, follow. And being wrong about the very rules that you are following is different from being wrong about some external matter. For, given that you're actually following the correct rules, you have access to very direct evidence about which rules those are. A thinker who is wrong about what rules they are following is wrong in the face of direct evidence, and this is irrational.

As Field says, though, this is a very problematic conclusion. Priest, Dummett, Rumfitt and other defenders of particular logics have spent decades making detailed arguments in favour of their proposed rules; this kind of activity seems like a paradigm of rationality. So if my view entails that all of them (perhaps excepting those who picked the right logic) are 'automatically irrational', then that is a problem for my view.

In fact, I don't think my view entails this. In particular, it's not true that following logical rules gives you direct evidence about which rules you are following. It's a familiar phenomenon that human beings can know how to do some rule-governed activity without being able to state the rule that they are following. This is often called 'tacit knowledge'; although merely labelling the phenomenon doesn't explain it, I think the familiarity of the phenomenon is enough to show that it is possible. The most prominent example of tacit knowledge is our knowledge of language (discussed below in 3.5.1). A competent English speaker can follow the kinds of rules that

are specified in a formal grammar of the language, but they need not be able to make those rules explicit. There is nothing puzzling in the example of a competent English speaker who has false beliefs about the rules of English grammar. This is the position of the vast majority of English speakers. So too for logic: knowing how to think does not require being able to state the rules you are following in thinking.

It might be argued that, even if thinkers need not be able to immediately state the rules they are following, they should be able to state these rules upon reflection. The rules may be known tacitly, but they should not be so deeply tacit as to be inaccessible. If they are inaccessible, then they aren't really 'known' or 'followed' at all: it is simply the case that the subject is disposed to conform to them, and we shouldn't use the label 'knowledge' for mere behavioural dispositions.⁶ But I don't think this is right: as I will argue in Chapter 3, the capacities involved in following logical rules have a first-personal aspect as well as a third-personal aspect. There is something it is like to follow a logical rule. So our tacit knowledge does not consist solely in dispositions to conform. But the first-personal aspect shows up in experiencing *particular* inferences as right or wrong; it need not put us in a position to know, merely by reflection, which rules we are following.

Against this, Zalabardo (2011, 134) argues that our capacity to recognize particular inferences as valid 'can be applied, not only to particular inferences, but also to types of inference'. This would mean that, if thinkers are able to recognize particular inferences as valid, they should also be able to recognize universal logical rules as valid. Zalabardo draws a comparison with grammar (2011, 134-135):

The very same ability that enables any English speaker to recognize the sentence *Mary arrived the house* as ungrammatical will enable speakers with the requisite concepts to acknowledge the truth of the proposition that the verb *to arrive* never takes a direct object. In this way, the capacity to recognize grammatical sentences can be a source of knowledge, not only of particular propositions, but also of universal generalizations.

But this seems implausible. The ungrammatical sentence sticks out as wrong, without any analysis. By contrast, to see that 'to arrive' never takes a direct object, we need to imaginatively consider a series of sentences – some with a direct object, others with an indirect object – and note which ones seem right and which seem wrong; then we can generalize from these instances to a rule, but our generalization might be wrong if we started with too limited a set of instances. So too for logic:

⁶Compare Dickie's objection to 'profoundly implicit' propositional attitudes (2017, 2582).

particular inferences seem right or wrong, but to get to a general rule, we need to generalize, and our generalization could be wrong if we started with too limited a set of inferences.

Indeed, many proponents of nonclassical logics might offer just this explanation of the general acceptance of classical logic. After all, these proponents often accept that classical logic is valid in many ordinary contexts (Russell 2019, 558). So, for example, an intuitionist who held that classical principles are invalid only in infinite settings (Brouwer 1908) might argue that our acceptance of classical principles is a result of generalizing wrongly from finite settings.

Moreover, when we experience a particular inference as right or wrong, it need not be clear whether the inference is right or wrong in virtue of a *logical* rule or some other rule. This is why logic students have to learn the difference between a materially valid inference like ‘If Moose Jaw is north of Timmins, then Timmins is south of Moose Jaw’ and a logically valid inference like ‘If Moose Jaw is north of Timmins, then something is north of Timmins’. This lack of direct first-personal access to the rules which make our inferences compelling provides another explanation for the existence of conflicting beliefs about which logic is correct.

So advocating a theory at odds with your practice isn’t always irrational. But I don’t mean to say that it’s *never* irrational. While following a rule doesn’t itself generate evidence about the rule you’re following, a third party might observe which rule you tend to conform to and point this out. If you have it pointed out to you that you’re affirming a rule which you don’t follow, or rejecting a rule which you do follow, you’re irrational if you don’t respond in some way. Suppose, for example, that you advocate intuitionistic logic and that someone points out that you actually follow Double Negation Elimination. There are several ways you might respond. First, you might take it as a reason to revise your theory and come to accept classical logic. Second, you might classify your conformity to DNE as an error and try to avoid making such errors in future. Third, you might give an alternative explanation – for example, that the cases where you followed DNE were all cases where you knew the thought was either true or false. More generally, you might try to explain your reasoning as materially rather than formally valid. If you don’t do any of these things, then you are irrational.

To conclude, a constitutivist has no trouble accommodating Type A disagreement: it doesn’t entail that the advocate of an incorrect logic is automatically irrational, let alone that they are failing to think.

2.2.2 Type B disagreement: using conflicting logics

Type B, the more difficult kind of disagreement to make sense of, occurs where two thinkers actually follow different logical rules. Suppose that R reasons classically, while D reasons intuitionistically, in the sense that R tends to accept all classical inferences and D tends to accept only intuitionistic inferences. If classical logic is correct, then R is following the correct rules and D is following the wrong rules. Are we forced to say that D is not thinking?

One possibility is that, deep down, D is really using classical logic but systematically refusing to draw certain inferences. Perhaps D has dispositions to draw classical inferences, but those dispositions are, in certain cases, blocked by D's theoretical commitment to intuitionism. It's a familiar fact that dispositions can be blocked or 'masked' in this way (Choi & Fara 2018). In this case, we could say that D is still thinking, although less well than R.

There are a couple of problems with this explanation. First, it is made plausible in the present case by the fact that intuitionism is weaker than classical logic, in the sense that if an inference is intuitionistically valid then it is also classically valid. This kind of explanation is less workable where the deviant thinker is using a logic stronger than the correct logic – for instance, if intuitionistic logic were correct, and classical logic the deviation. Of course, we could insist that the deviant thinker really has a set of dispositions corresponding to the correct logic, and that their theoretical commitments merely lead them to accept certain additional inferences which don't follow from the rules that their dispositions pick out, but this seems like an unmotivated position to take: it might be that the deviant thinker is genuinely disposed to draw the inferences specified by the incorrect logic.

Second, this explanation fails to take seriously the first-personal dimension of responsiveness to logical rules – the fact that the subject who has the capacity finds compelling inferences that fall under the rule. Given this fact, whether a subject's capacities correspond to classical or intuitionistic logic cannot be determined merely third-personally; it also depends on which inferences the subject finds compelling. In some cases, it might be that D really does find the classical inferences compelling. But if they only find intuitionistic inferences compelling, it would be wrong to say that they are 'really' following classical rules.

In cases where the above strategy is unavailable, and a subject does have capacities to follow logical rules other than the correct ones, we might not be forced to hold

that this subject is failing to think. We could instead loosen our view of what relation to logical rules is required for thinking. I argue in Chapter 3 that this relation involves logical capacities, but we might instead accept a view similar to that of Quine (1960) and Davidson (1973; 1985), requiring only that the subject tend to conform to the right logical rules, not that the subject have dispositions or capacities picking out those rules. In short: as long as you mostly conform, you are thinking, no matter what the explanation for your conformity is. A subject who tends to reason intuitionistically or relevantly or in accord with any other logic that largely overlaps with classical logic will tend to conform to classical logic.⁷ So on the Quinean view, such a subject will still be thinking. (The Quinean view does, however, allow for subjects who think ‘accidentally’: who, as a fluke, conform to the correct rules enough to count as thinking.)

If we don’t want to weaken our position in this way, then we are forced to hold that the deviant subject is failing to think. But this position, while hard to accept, is not as extreme as it might sound. For one thing, if our capacities can change over time, it may be that much of the subject’s activity still counts as thinking: only those segments of the activity during which the subject’s capacities pick out an incorrect logic fail to count as thinking. For another thing, as I’ll argue in Chapter 5, our failures to think can be subjectively indistinguishable from thinking, and can involve reasoning and the use of that-clauses. So the claim that someone is failing to think is not refuted by showing that their activity looks like thinking, or that they are under the impression that they are thinking.

Finally, I want to emphasize that this anti-deviant conclusion is premised on accepting that there is one true logic. It shouldn’t be surprising that, for a monist, those subjects who follow deviant logics are at risk of producing nonsense. Conversely, much of our dissatisfaction with this conclusion is, I think, really a dissatisfaction with the idea that one logic is absolutely correct. The felt need to ‘save’ deviant thinkers from the conclusion that they fail to think really reflects an implicit sympathy for pluralism. This takes us to the next stage of the discussion.

⁷We might even weaken the position still further to hold that which logical rules you count as following is determined in part by your environment, along the lines of the externalism about logical form discussed in Ludlow (2003). But I find this degree of externalism hard to make sense of.

2.3 Pluralism and Disagreement

At this stage we drop the monist assumption that there is a single correct logic, allowing instead that more than one logic is correct – in particular, that there may be multiple correct (and non-equivalent) sets of rules for the conditional and negation. Dropping this assumption leaves the claim that logic is constitutive of thinking ambiguous. Once we allow that more than one logic is correct, it's no longer clear which logical rules a thinker must tend to conform to.

One response would be to allow that following any finite set of logical rules is sufficient (other things being equal) for thinking. This position tallies well with Warren's (2015, 4) 'unrestricted inferentialism', the view that any set of inference rules for an expression suffices to determine a meaning. Warren concludes that a language containing Prior's *tonk* is a possible language, whose sentences express propositions.⁸ This response would be consistent with logical pluralism, but it would make the constitutive claim nearly vacuous: any subject who followed some set of logical rules would count as thinking.

I am going to argue that constitutivism is consistent with two more limited kinds of logical pluralism. The first kind, which I call *domain variance*, holds that there are multiple domains of thoughts, such that while there's a single correct logic for each domain, different domains can have different logics. The second kind, which I call the *minimal kit*, holds that for a given domain of thoughts there are multiple correct logics, such that following any one of them is sufficient for thinking. Both responses, I will argue, allow for a degree of logical pluralism while maintaining the core of the claim that logic is constitutive of thinking. I'll suggest, however, that the first response is feasible only if the second is.

2.3.1 Domain variance

In an early paper called 'Truth', Dummett argued that we are entitled to reason classically with statements in a given domain if, but only if, statements in that domain are decidable (1959, 66). When it comes to domains where statements are not decidable, we ought to reason intuitionistically instead. Dummett's claim is

⁸Here are the rules for *tonk* (Prior 1960).

Tonk Introduction From p , infer $p \text{ tonk } q$.

Tonk Elimination From $p \text{ tonk } q$, infer q .

an instance of *domain variance*: the view that different logics might be correct for different domains of thought.

This idea is taken up by Kissel & Shapiro (2017, 3), who argue that ‘the correct logic is dependent on the domain of discourse’. Formal logic spells out the implicit norms of thinking, but these norms may be different depending on what we’re thinking about. Even within mathematics, Kissel & Shapiro argue, some theories are consistent only on the assumption of a logic weaker than classical logic.⁹ So

the norms of various logics are constitutive of various mathematical theories: there are classical theories, intuitionistic theories, and perhaps even relevant theories, quantum theories, sub-structural theories, ... If you are working in a classical theory, you ought to reason in such and such a way; if in an intuitionistic theory, in this and so way, etc. (2017, 17)

The upshot: while there’s a single correct logic for reasoning in any given domain, there’s no single correct logic for reasoning as such (2017, 20):

the reason why there is no One True Logic is that the norms implicit in these practices are different from each other. Each of the logics is truth-preserving in its domain, in the sense that if the premises of valid argument (in the indicated logic) are true in the domain, then so is its conclusion. None of the logics, or perhaps only the weakest of them, is truth preserving across the board.

Though they don’t discuss this, Kissel and Shapiro’s pluralism seems equally applicable to domains outside of mathematics. For example, there may be different logics for thought about ordinary objects, ethical thought, mathematical thought and scientific thought.

Does domain variance provide a way to make constitutivism consistent with logical pluralism? A dilemma posed by Haack suggests not. Haack argues that

if the ground for distinguishing the domains refers to content, one will be disposed to say that the systems are not really logical, and if the ground for distinguishing the domains refers to form, one will be disposed to say that the systems are not really rivals. (1996, 46)

⁹For example, Smooth Infinitesimal Analysis (a theory of infinitesimals) is inconsistent if the background logic is classical, but consistent in intuitionistic logic (Shapiro 2014, 72-75). There are serious questions about whether Smooth Infinitesimal Analysis makes sense, even granting that it is consistent (Rumfitt 2018, 15), but other examples are available: for example, naive set theory with a dialetheist background logic. A different sort of example involves theories that are distinct only when the background logic is distinct, such as Heyting Arithmetic and Peano Arithmetic.

Let's take the horns of the dilemma in turn.

On the one hand, if the domains are distinguished by the logical form of the thoughts involved – for example, if one logic is appropriate for tensed thoughts and another for untensed thoughts, or one logic for thoughts containing modal operators and another for thoughts without them – then domain variance is not really a kind of pluralism. It's equivalent to Hjortland's proposal that we only need 'one logical theory, but the theory itself recommends restricted logical principles for different parts of the language' (2017, 654). For we can just conjoin the logics to get a single correct logic for all domains. If this is pluralism, then anyone who accepts both propositional and predicate logic is a pluralist.

On the other hand, Haack argues, if the domains are distinguished by subject-matter, then the rules that vary between domains are not topic-neutral; if logical rules are topic-neutral, then the rules that vary between domains are not logical. So we have pluralism, but not logical pluralism. (Moreover, if only the rules which hold in all domains are logical, then there's a risk of logical nihilism (Russell 2018b, 345), or at least a very weak logic.)

I'm going to argue that we should resist this horn of the dilemma. Before I do so, however, let me strengthen the problem by noting an additional way that domain variance seems to put pressure on topic-neutrality. Consider the question of what logic holds for thoughts that cross different domains. For example, if p is a mathematical thought and subject to intuitionistic logic, while q is an ethical thought and subject to classical logic, what is the right logic for p and q ?

While this example might seem artificial, the fact that domains of thought are not completely sealed off from each other seems to guarantee that such domain-crossing thoughts will sometimes occur. One possible response is that where the logic for p is weaker than the logic for q , the weaker logic holds for complex thoughts composed of p and q . But there is not always a weaker logic: sometimes the sets of entailments recognized by two logics will overlap partially, with neither being contained in the other.

We might hold instead that there's a minimal background logic that governs domain-crossing thoughts – perhaps the intersection of the logics for the different domains. But now there's some pressure – derived, again, from the thought that logic is topic-neutral – to say that only the minimal background logic is really logic, and the rules for particular domains just reflect general principles about those domains. In other words, the pluralist has to resist the proposal, made by some proponents of

nonclassical logics (Beall 2018), that a weak logic specifies the core of consequence relations in all domains, and in each domain additional (non-logical) principles hold which strengthen the local consequence relation. The proponent of the weak logic ‘recaptures’ stronger rules locally without taking them to be part of the logic.

Both Haack’s dilemma and the domain-crossing problem seem to show that domain variance is not a kind of *logical* pluralism, because the domain-variant rules violate topic-neutrality. If these arguments are right, they cast doubt on the consistency of logical pluralism with constitutivism. One way out might be to abandon topic-neutrality, but that way isn’t open to me. As I argued in Chapter 1, the thought that logic is topic-neutral, in the sense that logical concepts are always applicable, is not merely traditional: it’s connected with the expressive role of logical concepts, which constitute a general, open-ended ability to respond to tension and support among thoughts. My only way out, then, is to show that this sense of topic-neutrality is consistent with domain variance.

Here’s the idea. While the precise rules for negation may vary from domain to domain, there’s a single, topic-neutral notion which is at the core of negation in any domain. Intuitively, while the early Dummett thought that Double Negation Elimination was appropriate in some domains and not others, he took negation itself to have a common meaning across domains. In every domain, negation expresses tension between thoughts. And the same might be said about the conditional: while the precise rules may vary, in every domain it expresses support between thoughts. One way to put this would be that the rules in different domains constitute different *conceptions* of a single *concept*.

If this idea can be made to work, then the concepts of negation and the conditional are topic-neutral, while the precise rules for these concepts can vary with subject-matter, as required by domain variance. In this way, domain-variance is consistent with topic-neutrality, and so pluralism is consistent with constitutivism.

To make this idea work, we need to spell out the sense in which the logical concepts can remain the same while some of the rules vary. What does the core of a logical concept consist in? One approach to this problem would be to find common meanings for the logical constants which hold in different logics, against Quine’s (1970, 100) claim that a change of logic is a change of meaning.¹⁰ Here I’m going to

¹⁰For model-theoretic approaches, see McDowell (1997a), Rumfitt (2012) and Fine (2014). For proof-theoretic approaches, see Restall (2014), Hjortland (2013) and Dicher (2016). Quine later takes a related position, holding that some logical truths are analytic – in the attenuated sense that we learn their truth in learning the words involved – while others are not (1974, 80).

defend only a necessary, not a sufficient condition for a logic to contain negation and the conditional. I'm going to argue that such a logic has to include certain *minimal principles*.

Note, however, that once we have this idea, we can also use it to develop a second kind of logical pluralism. There might be only a single domain of thoughts, but multiple different logics suitable for thinking about the domain because they all contain the same logical concepts. I develop this idea in the next section.

2.3.2 The minimal kit

I've argued that constitutivism is consistent with one kind of logical pluralism – domain variance. In this section, I develop a second kind of pluralism, on which for a given domain of thoughts, there is more than one correct logic, such that following any correct logic is sufficient (other things being equal) for thinking. I argue that this kind of pluralism is consistent with constitutivism. It's also consistent with domain variance, but for simplicity I'm going to assume, in this section, that there is only one domain of thoughts.

The proposal is that there is a (limited) range of logics such that a subject following any logic in the range can count as thinking. Any logic in the range has to contain certain rules, which I call the *minimal kit* (Hale 1999, 51; 2002; Leech 2015). I'll begin by discussing which rules a logic must contain to be suitable for thinking with. Following this, I consider an objection about the sense in which logic is constitutive of thinking on this proposal. Finally, I make some brief comments on which deductive systems contain the minimal kit.

So: which rules are in the minimal kit? One proposal comes from recent work by Suki Finn (2019a, 2019b). Finn argues that there are certain logical rules which a logic must include if it is to be adequate. These rules have the property of being 'self-governing': they are of the very structure that they govern. For example, Finn argues that all rules of inference are conditional in structure: they are of the structure *If the premises instantiate structure X, then infer conclusion Y* (2019b, s. 3). She then argues that Modus Ponens governs all conditionals, and therefore governs all rules of inference. It's not entirely clear what is meant by 'governs', but a plausible reading is that a rule governs a structure iff sensitivity to the rule is required for understanding instances of the structure. In the present context, the idea would be that to understand conditional structures you have to follow Modus Ponens. This reading echoes Boghossian's suggestion that 'there is no alternative but to accept

“conditional theory”—Modus Ponens in effect—if you are so much as to have the conditional concept’ (2003, 247). Now, given that Modus Ponens is itself a rule of inference, it is itself of conditional structure, and is thus self-governing in Finn’s sense.

An initial proposal, then, might be that the rules in the minimal kit are precisely those that are self-governing. But while I do want to build on Finn’s ideas in giving an account of the minimal kit, I don’t want to use them so directly, for several reasons. First, very few rules are self-governing in Finn’s sense: she proposes Modus Ponens and Universal Instantiation, and rejects other contenders. But requiring only these rules would impose a very weak constraint on the range of possible logics for thinking. Second, Finn’s view of self-governance relies on the assumption that all logical rules are conditional in form. But it’s not clear that this is the case. Some logical rules, like the Law of Non-Contradiction, are, intuitively, not even broadly conditional in structure. The Law of Non-Contradiction could be formulated as *it is not the case that p and $\text{not-}p$* , or *do not judge p and $\text{not-}p$* , or *it is incorrect to judge p and $\text{not-}p$* ; none of these are governed by Modus Ponens.

The third point is not an objection to Finn’s notion of self-governance, but to the idea of using it to determine which rules are in the minimal kit. It’s simply unclear why self-governance should be a condition on being constitutive of thinking. Finn argues that self-governing rules generate the *adoption problem* – the problem that it would be impossible for someone who didn’t already follow the rule to adopt it, because adopting it presupposes that you already follow it (Padró 2015). This is compelling; but generating the adoption problem doesn’t seem like a sufficient condition for being constitutive of thinking. There could be a rule that is unadoptable – perhaps it is of some alien and self-governing structure – but which is not constitutive of thinking.

In fact, it seems to me that self-governance is not the most fundamental notion in Finn’s account. A rule’s being self-governing is a consequence of the fact that the rule governs all rules of inference. For example, Modus Ponens is self-governing because it governs all inferential rules, or at least all rules of conditional structure. This suggests that we should look more directly at which rules govern all inferential rules, or at least which rules govern the general structures that rules have: if there are any such, they’ll have a strong claim to being constitutive of thinking. But are there?

It’s helpful to start a little bit further back, by considering the types of rational

relation in which thoughts stand to each other. By ‘rational relation’, I mean a relation that plays a role in fixing a thought’s place in the web of thoughts. I argued in 1.2.1 that there are at least two general kinds of rational relations: *support* and *tension*. A thought might support another thought, as *this is an oak* supports *this is a tree*. And a thought might stand in tension with another thought, as *this is an oak* stands in tension with *this is a maple*. I argued that appreciating such rational relations requires grasp of the conditional and negation. If there are core rules for the conditional and negation, these will be the rules in the minimal kit. (The same rules will also be self-governing in roughly Finn’s sense: just as any codification of the rules for the conditional will presuppose some notion of support, any codification of the rules for negation will presuppose some notion of tension.)

Which logical rules, then, are at the core of the conditional and negation? Two plausible candidates, which I proposed in 1.4.1, are Modus Ponens and Non-Contradiction. Appreciating that p supports q requires sensitivity to Modus Ponens, which says, in intuitive terms, to believe q if you believe p and you take it that p supports q . Appreciating that p and r are in tension requires sensitivity to Non-Contradiction, which says not to believe p and *not-p*. So these rules are such that a thinker must be sensitive to them in order to appreciate any relation among thoughts, and so in order to understand any thoughts at all. I claim, then, that a logic adequate for thinking with must contain Modus Ponens and Non-Contradiction.¹¹ Just how to formulate these rules is discussed in the next section.

Note, however, that these are likely not the only rules that are required. For one thing, Modus Ponens is a rule for eliminating the conditional; a usable logic would also need some way of introducing the conditional. Perhaps this:

Conditional Proof If $\Gamma, P \vdash Q$, then $\Gamma \vdash P \rightarrow Q$

We might also need rules for other connectives, such as conjunction or disjunction.¹²

¹¹Lukasiewicz held a similar view and even proposed the same kit, plus a rule of substitution (1937, 248): ‘Absolute truths of thought did not collapse in 1930. Whatever discredit anyone may try to cast upon many-valued logics, he cannot deny that their existence has not invalidated the principle of exclusive contradiction. ... Also valid remain the rules of inference, namely the rule of substitution, which corresponds to the Aristotelian *dictum de omni*, and the rule of detachment, analogous to the Stoic syllogism called *modus ponens*. Owing precisely to these rules we are building today not one but many logical systems, each of which is consistent and free of contradiction. It may be that other absolute principles, with which all logical systems must comply, also exist.’

¹²Maddy (2002, 71) argues that the rudimentary logic we think with uses conjunction and disjunction with Strong Kleene truth-tables, and the conditional and negation with partial truth-tables; in (2014, 94) she accepts the Strong Kleene tables for the conditional and negation too.

For another thing, I haven't said anything about structural rules – rules which are closed under substitution for sentences (Restall 2000, 24). It seems likely that Reflexivity is in the minimal kit:

Reflexivity $P \vdash P$

But other structural rules, like Weakening, are more controversial (Hlobil 2016, 90; Brandom 2018, 72).

Weakening If $\Gamma \vdash Q$, then $\Gamma, P \vdash Q$

So the minimal kit will need some structural rules, but I haven't said which. A thinker might also need a rule of substitution to get from general rules to instances of those rules (Besson 2019a, 184; Cohnitz & Nicolai (ms.)). Finally, the minimal kit might need some rules governing subsentential structure, such as Universal Instantiation.¹³ In short: my claim is that MP and LNC are *in* the minimal kit, not that they exhaust it.

A natural objection arises at this point: isn't the minimal kit the One True Logic? If it is, then the minimal kit proposal is not a kind of pluralism at all, but a minimalist kind of monism. But I don't agree that the minimal kit is the One True Logic. The reason for this is that a subject who followed only the rules in the kit would not be able to think: the rules are just too minimal. A logic suitable for thinking needs more than just the minimal kit; the pluralist point is that there's no unique *more* that it needs. There are many different ways of supplementing the kit to get a logic suitable for thinking. (Of course, even following a logic suitable for thinking is not *sufficient* for thinking – as I emphasized in Chapter 1, there are other necessary conditions which I haven't investigated.)

This response invites a second objection, which parallels the objection against domain variance we discussed earlier. Consider some rule – say, Double Negation Elimination – which is not in the minimal kit, but which is in some of the logics suitable for thinking. It's possible to think without following DNE, because a thinker could follow intuitionistic logic, which doesn't contain it. So DNE is not constitutive

¹³Brandom (1994) tries to prove that any language with negation and a conditional must have subject-predicate structure if it has subsentential structure at all. But the argument is very hard to pin down. Moreover, subsentential structure raises novel issues. There's a question about how logical form relates to empirical syntactic structure, and whether it is a problem if empirical syntax turns out to have no use for categories like *singular term* and *predicate* (Collins 2007; 2015).

of thinking. But (given what I've argued) logical rules have to be constitutive of thinking. So DNE is not a logical rule. In general, then, rules outside of the minimal kit are not logical rules; they are contingent, not necessary for thinking. And this means that, while the minimal kit might be a kind of pluralism, it's not *logical* pluralism.

This objection brings out an important point. Accommodating pluralism along these lines does force us to revise the sense in which logic is constitutive of thinking. Previously, we said that logical rules are constitutive of thinking in that subjects have to follow them in order to think – but this excludes the rules outside the minimal kit. Instead, we should say that logical rules are constitutive of thinking in that subjects have to follow *one of the right logics* in order to think. This is a version of the 'relativized constitutivism' which Steinberger (2017, 156) attributes to Carnap:

Might it be constitutive of thought not that there is a unique set of logical norms to which thinking is subject, but rather that the thinker acknowledge the normative authority of one set of logical laws among a range of different but equally acceptable options? According to this picture there are more than one, potentially numerous, sets of norms. None of them has an exclusive claim to being constitutive for thought, but in order to count as thinking one must acknowledge the authority of exactly one (any one) of them. ... On this view, thought, much like the rule of law, is characterized by its *multiple normative realizability*.

Carnap's view requires acknowledging the authority of the rules, while mine requires actually following them, but the structure of the views is the same. To think, you need a logical framework, but no framework in particular. Given this modification, we can hold that DNE is constitutive of thinking, relative to logics which contain it, and therefore that it is a logical rule. A subject who follows classical logic has to follow DNE in order to think (unless they adopt a different logic), but not every subject needs to follow classical logic.

2.3.3 Recap

The arguments of this section have been a bit tangled, so let me sum up my position. Constitutivism is consistent with two kinds of pluralism: different domains can have different correct logics, and a single domain can also have multiple correct logics. Putting these kinds together, we have a picture in which subjects can differ in the rules they follow both within and between domains, while still thinking.

To illustrate: suppose there are two domains (the finite and the infinite) and two logics (classical and intuitionistic). Then there are four ways that a subject might think: classically in both domains; intuitionistically in both domains; classically for the finite and intuitionistically for the infinite; or the reverse. Now, if we consider the range of possible domains and the range of possible logics, it's clear that there are many more possibilities than this.

To be clear, however, I'm not saying that a subject who followed any of these possibilities would be *right*, all things considered. But they wouldn't be failing to think, at least not for logical reasons.

The resulting position is both pluralist and constitutivist. Different individuals, different communities or different species of thinkers might follow different rules, in one domain or in all of them, but it's still not the case that anything goes. For example, we deny that a community might follow the rules for the trivializing connective *tonk*, *pace* Warren (2015). In Warren's interpretation, the Tonkers have no negation or conditional in their language (nor any other non-trivial connective): for example, the Tonkish symbol equiform with our 'not' expresses a vacuously unary constant function which takes any value to *true*.¹⁴ We also rule out less extreme logics. For example, we also rule out Kripke's (perhaps rather tongue-in-cheek) proposal that a positive logic, with no connective that takes truth to falsity, is 'adequate for science' (2015, 383; cf. Brîncuș & Toader 2018).

2.4 Which logics contain MP and LNC?

Which logics contain the rules in the minimal kit? It turns out that this is not a straightforward question, particularly when it comes to the Law of Non-Contradiction. So I won't attempt a definitive answer here, but will merely make some tentative remarks. In this section I will use 'Modus Ponens' and 'Law of Non-Contradiction' to refer to the rules for thinking which I motivated above and in Chapter 1:

Modus Ponens From p and *if p , then q* , infer q .

Law of Non-Contradiction Do not judge p and *not- p* .

I'll use different terms to refer to the rules which hold, or fail to, in various deductive

¹⁴The Tonkers can nonetheless follow standard classical rules for negation only because those rules don't fix the standard semantics; as Warren observes (2015, 18), the Tonkers wouldn't be able to follow the bilateral rules I'll discuss at the end of the next section.

systems, and which aim to capture the force of Modus Ponens and the Law of Non-Contradiction.

As above, I refer to a logic as a set of entailments between sentences of an artificial object-language together with some *superstructure*, a model theory or proof theory, which generates the entailments. An entailment is a metalanguage sentence of the form $\Gamma \vdash Q$, where Γ is a (possibly empty) set of sentences. This builds in a few assumptions which I want to flag. First, I'm assuming that an entailment can have only a single conclusion. Some theorists argue for multiple-conclusion logics, where an entailment is of the form $\Gamma \vdash \Delta$, and says that if all sentences in Γ are true, at least one sentence in Δ is true (Shoemith & Smiley 1978). In principle, I don't want to rule out the possibility that some subject might think by following a multiple-conclusion logic. But I won't consider the possibility here; in any event, multiple-conclusion arguments don't seem to be part of *our* reasoning practice (Steinberger 2011b).

Second, I'm assuming that Γ is a *set* of sentences, not a *sequence*. In other words, the order of premises doesn't matter, and repetition of premises doesn't matter (Restall 2000, 3). This amounts to assuming two structural rules:

Exchange If $\Gamma, P, R \vdash Q$ then $\Gamma, R, P \vdash Q$

Contraction If $\Gamma, P, P \vdash Q$ then $\Gamma, P \vdash Q$

Again, both of these rules might be denied. If we suppose that premise combination stands for concatenation, as in categorial grammar (Lambek 1958), Exchange fails. In a resource-conscious logic where premises get 'used up', like linear logic, Contraction fails (Troelstra 1992). Someone might make a case that the correct logic for thinking should reject either or both of these, but I won't have room to consider that possibility here.¹⁵

Third, while it's fairly clear how a proof theory generates an entailment, there are more options in the model-theoretic case. For present purposes, $\Gamma \vdash Q$ means that whenever all of Γ have a designated value, Q has a designated value. The designated values are *true* and (in logics that allow it) *both true and false*. But you might also define entailment as preservation of only truth, or preservation of non-falsity. Different definitions will generate different sets of entailments from the same truth-tables.

¹⁵Beall & Murzi (2013, 166) and Shapiro (2015, 80) propose dropping Contraction to block a version of Curry's paradox.

Now I can set out the logics I will consider. I'm only going to consider propositional logics, and I won't fully lay any of them out; since the focus is on which logics contain the philosophically motivated rules above, I'll only refer to some salient features of each logic, providing references to places where full presentations can be found.

- Classical logic or **CL**, with the connectives interpreted in the usual way using two-valued truth-tables, such that every proposition is either true or false, and no proposition is both. For defences of classical logic, see Rumfitt (2015) and Williamson (2018).
- Intuitionistic logic or **IL**, which rejects Double Negation Elimination ($\neg\neg P \vdash P$) and the Law of Excluded Middle ($\vdash P \vee \neg P$). For arguments in favour of intuitionism based on objections to classical reasoning about infinities, see Brouwer (1908; 1951) and for arguments based on the theory of meaning, see Dummett (1991). For a presentation see van Dalen (2001).
- The relevant logic **R**, which, like other relevant logics, aims to respect the principle that in a valid argument the premises should be relevant to the conclusion, so as to avoid 'paradoxes of material implication' such as $P \rightarrow (Q \rightarrow P)$. R is presented in Anderson & Belnap (1975).
- Strong Kleene Logic or **K3**. Kleene (1952, §64) developed three-valued truth-tables for the logical connectives, such that a sentence can be true, false or *neither*. The third value is undesignated. It might apply in cases of indeterminacy of truth-value, ignorance of the correct truth-value, reference failure or some other defect. The entailments of K3 are those which preserve truth given this semantics (importantly, this isn't the same as preserving non-falsity).¹⁶
- Graham Priest's Logic of Paradox **LP** (Priest 2006), which is also three-valued, but the values are true, false and *both true and false*. The truth-tables for LP are the same as those for K3, but the third value represents both rather than neither. So LP has two designated values. The entailments of LP are those which preserve either true or both: in other words, if all of Γ are either true or both, Q is either true or both.

¹⁶There are other truth-tables with two undesignated values, such as Ł3 (Łukasiewicz 1930) and the 'internal' B3 (Bochvar 1937), also known as Weak Kleene Logic, which I won't consider here. These truth-tables and others can be found in Haack (1996, 168).

- The logic of First Degree Entailment or **FDE**, which is four-valued, allowing sentences to be true, false, both or neither (Dunn 1976). The entailments of FDE are those which preserve designation, i.e. either truth or both, like those of LP. FDE has a generality which is sometimes used to argue that it is the correct logic (Beall 2018): if you take FDE and then rule out sentences taking the value neither, you get LP; if you rule out the both value, you get K3; and if you rule out neither and both, you get CL. For discussion and defence of FDE, see Anderson & Belnap (1975, §81) and Beall (2018).

I don't at all claim that this exhausts the main contenders for a logic to think with, but it's a reasonably broad selection which will give us the lay of the land.

Let's start with Modus Ponens. There are a couple things we might mean in asking whether a given deductive system 'contains' Modus Ponens. First, we might be asking whether

Detachment $P \rightarrow Q, P \vdash Q$

is a valid form of argument within the logic. On a model-theoretic approach, Detachment is valid iff Q is true (or: has a designated value) on condition that P and $P \rightarrow Q$ are true (or: have designated values). On a proof-theoretic approach, Detachment is valid iff Q can be proven from $P \rightarrow Q$ and P .

Second, we might be asking whether the following statement, known as Pseudo Modus Ponens, is valid:

Pseudo Modus Ponens $\vdash (P \wedge (P \rightarrow Q)) \rightarrow Q$

Now, in the present case, I think it's clear that Detachment captures our informal idea of Modus Ponens better than Pseudo Modus Ponens does. As I argued in 1.5, a logic that didn't contain any rules of inference (like Detachment) but only the conditionals that corresponded to them (like Pseudo Modus Ponens) would be unable to treat any proposition as supporting any other proposition. Still, considering the standing of Pseudo Modus Ponens within different logics will help set up our discussion of the Law of Non-Contradiction below.

So let's ask: which logics validate Modus Ponens?

Detachment is valid in CL and IL and in the relevant logic R. Pseudo Modus Ponens also holds in these logics.

Detachment also holds in the three-valued logic K3. LP and FDE lack a conditional; if we define $P \rightarrow Q$ in the usual way, as an abbreviation for $\neg(P \wedge \neg Q)$, then

Detachment fails for this conditional. Pseudo Modus Ponens holds in LP, which validates all classical theorems, but it fails in K3 and FDE, which in fact have no theorems. Feferman (1984, 264) comments that, in light of these problems with the conditional, ‘*nothing like sustained ordinary reasoning can be carried on*’ in K3 or LP, and Hazen & Pelletier (2020) say the same about FDE.

But there is a lot more to say about this. In particular, it is possible to *add* a detachable conditional which is not defined in terms of negation and conjunction. Hazen & Pelletier (2020) propose adding such a conditional to FDE, K3 and LP. So these logics with a conditional added would, it seems, be able to express relations of support. However, the conditional of Hazen & Pelletier violates other familiar principles, including Contraposition, $P \rightarrow Q \vdash \neg Q \rightarrow \neg P$. We might be willing to abandon Contraposition, but this leads to a more general point.

In asking whether Detachment holds in various logics, we have taken for granted some conception of what makes the \rightarrow a conditional in these logics, beyond merely being subject to Detachment. Being subject to Detachment might be necessary for being a conditional, but it can’t be sufficient. If all we required was that there be some connective \otimes , such that $P \otimes Q, P \vdash Q$, then any logic containing conjunction would meet the test (since if $P \wedge Q \vdash Q$ then, by Weakening, $P \wedge Q, P \vdash Q$). But conjunction is not a conditional. So more is required to count as a conditional than satisfying Modus Ponens. *What* more is a hard question, disputed by proponents of different logics, and not something I’ll try to settle here.

For present purposes, then, my conclusion is that a subject who followed the rules of CL, IL, R or K3 would be able to express relations of support, and that a subject who followed the rules of FDE or LP plus a detachable conditional would too, presuming that the added connective was a genuine conditional.

Things get even less straightforward when we ask which formal logics contain the Law of Non-Contradiction. There are various things we might take this to mean. First, we might be asking whether the following sentence, which I’ll call the *Principle of Non-Contradiction*, is a logical truth or theorem:

Principle of Non-Contradiction (PNC) $\vdash \neg(P \wedge \neg P)$

CL, IL and R all validate PNC, while it fails in K3 and FDE. LP is a well-known tricky case (Grim 2004): PNC is a logical truth in LP, but it can also be false. This is possible because in LP a statement can be both true and false.

However, there’s some reason to doubt that PNC is the right way to formally

capture the Law of Non-Contradiction. For one thing, as LP shows, a logic can validate PNC while still tolerating contradictions, and even allowing them to be true. More philosophically, PNC is for Non-Contradiction what Pseudo Modus Ponens was for Modus Ponens: both condense a logical *rule* into a logical *truth*. Non-Contradiction, conceived as a rule, prohibits judging contradictions; PNC states that the negation of a contradiction is a logical truth, such that we are always entitled to judge it.

Before considering more radical proposals, let's consider a more familiar one. We might try to formally capture the Law of Non-Contradiction by asking whether, in the system, a contradiction entails everything:

Explosion $P, \neg P \vdash Q$

One line of reasoning in favour of Explosion goes like this. In an explosive logic, a contradiction entails everything. But we know that not everything is true. (This is reminiscent of Putnam's (1978) argument that 'not every statement is both true and false' is an unrevisable *a priori* truth.) In an explosive logic, then, there is very strong reason not to accept a contradiction. And this is a sense in which Explosion expresses some of the force of the Law of Non-Contradiction. By this standard, CL, IL and K3 are in; paraconsistent logics such as LP, FDE and R are out.¹⁷

But there are problems with this solution too. Intuitively, there's a difference between 'don't believe a contradiction' and 'from a contradiction, infer everything'. Given this difference, it's possible for a proponent of a nonclassical logic, such as a relevantist, to object to Explosion on the basis that P and $\neg P$ may have nothing to do with Q , without evincing any willingness to accept contradictions. Moreover, Kürbis (2019, 111) argues that 'it is a contingent feature of a language that some sentences are false'. Even if a language all of whose sentences were true would not be learnable, it would seem strange to rely on such considerations in our definition of a logical law.

Another possibility is that Non-Contradiction should be understood not as a rule or theorem, but as a semantic principle: in my earlier terminology, we should look for it in the superstructure of a logic rather than in the entailments. In particular, we might require that the logic satisfy the following principle:

Semantic Non-Contradiction (SNC) P and $\neg P$ are not both true.

¹⁷For a discussion of other forms of Non-Contradiction derivable in R, see Restall (2004).

If we maintain that it is correct to judge a proposition only if the proposition is true (or designated), then if Semantic Non-Contradiction is satisfied, it will never be correct to judge both P and $\neg P$. Among the logics we've considered, Semantic Non-Contradiction holds in CL, IL, R and K3; it fails in LP and FDE, both of which allow for P and $\neg P$ to be true together (as P is both true and false iff $\neg P$ is).

There is, however, an additional difficulty that arises if we want to say that thinkers should follow Semantic Non-Contradiction. As a principle about the truth-values of propositions, SNC is not the kind of thing that our logical capacities can pick out; unlike the other logical rules we've considered so far in this subsection, SNC doesn't license or mandate any particular pattern of inferences. (You might say that SNC rules out judging that P and $\neg P$ are both true: but this form of judgment, which is available only for subjects who can think about their own thoughts and predicate truth of them, is not what we were looking for. Non-Contradiction should constrain our relation to P and $\neg P$, not to P *is true* and $\neg P$ *is true*.)

This makes it unclear what it would be for a subject to follow SNC. The natural suggestion is that, even if a subject doesn't follow SNC directly, the logical rules that the subject follows should *force* SNC to hold. But this natural suggestion collides with the *categoricity* problem pointed out by Carnap (1943). The problem is that the standard classical entailments do not fix the standard interpretation of the logical connectives, and in particular do not rule out interpretations in which P and $\neg P$ are both true. To see this, consider that an entailment $\Gamma \vdash P$ holds in an interpretation iff it is not the case that all of Γ are true while P is false. All classical entailments, then, will hold in an interpretation in which all propositions are true. But in such an interpretation, P and $\neg P$ are both true.

What this means is that, if we want the logical rules that the subject follows to force SNC to hold, then those rules cannot be the standard classical rules. The problem is that, as Smiley (1996, 7) writes, consequence 'depends exclusively on the impossibility of the premisses being true and the conclusion false, and the possibility or otherwise of their all being true is irrelevant'. In other words, entailments can require that if Γ are true, P is true, or that P is always true, but they cannot require that if Γ are true, P is false, or that P is always false. This precludes them from forcing SNC to hold (Church 1944; Button & Walsh 2018, 303).¹⁸

Why is it that Modus Ponens was relatively straightforward to specify as an

¹⁸Raatakainen (2008) argues that the same problem applies to intuitionistic logic, although Murzi & Hjortland (2009) respond that intuitionists can avoid it by identifying truth with proof.

entailment, while Non-Contradiction seems so elusive? I think we can see the problem if we compare some intuitive characterizations of the Law of Non-Contradiction with the formal notions we have been using. Intuitively, Non-Contradiction says that ‘a proposition and its negation cannot both be accepted’ (Price 1990, 224). Along similar lines, Peacocke (1987, 163) writes that ‘[w]hat is primitively obvious to anyone who understands negation is just that $\neg A$ is incompatible with A ’. What we want, intuitively, is a rule of prohibitory form: a rule that precludes accepting, or mandates rejecting, contradictory contents.

But the standard conception of an entailment assumes that all rules are of conditional form: it says that *if* the premises are true, *then* the conclusion is true, or that the conclusion *can be* proven from the premises. As Kürbis (2019, 122) observes, standard proof theories use only

what might be called *positive* primitive notions: affirmation, assertion, truth. They avoid *negative* primitives, such as negation, denial, falsity. The latter are supposed to be *defined* in terms of the primitive notions of the theory. Put this way it may not be so surprising that the project fails: it is a natural thought that nothing negative can be got from the purely positive. The philosophical insight I will tentatively draw ... is that *negative* as well as *positive* primitive notions are needed in the theory of meaning.

Kürbis concludes that it is impossible to define negation proof-theoretically without using some negative primitive notion (see also Sylvan 1999, 305).

This, I think, is the source of our difficulties with the Law of Non-Contradiction. If the Law of Non-Contradiction is of prohibitory rather than conditional form, then we will need to introduce some negative primitive notion to adequately formalize it. Moreover, if Kürbis is right that defining negation requires a negative primitive, then negation is likely self-governing in Finn’s sense, because it is impossible to understand explicit rules for negation without already being sensitive to those rules.

Let me sketch one proposal that uses a negative primitive notion.¹⁹ Bilateralist theorists propose that we recognize *rejection* as a cognitive act on par with judgment or assertion. They argue, *pace* Frege (1919), that rejecting a thought is not reducible to judging or asserting its negation; the logical rules we follow specify not only what we should judge, but also what we should reject. In Rumfitt’s (2000) proposal, we extend

¹⁹Another proposal is to define negation in terms of incompatibility. Incompatibility accounts have been given for CL (Peacocke 1993) and intuitionistic relevant logic (Tennant 1999), but I’m not sure how to extend them to the other logics we’re considering.

the object-language with *force markers*: $+$ marks the force of judgment/assertion and $-$ marks the force of rejection. The rules governing negation, then, look like this:

$+\neg$ **Introduction** $-P \vdash +(\neg P)$

$+\neg$ **Elimination** $+(\neg P) \vdash -P$

$-\neg$ **Introduction** $+P \vdash -(\neg P)$

$-\neg$ **Elimination** $-(\neg P) \vdash +P$

We also have a reductio rule due to Smiley (1996; Rumfitt 2002, 312n). Where α, β range over propositions with force markers and α^* is the result of reversing the force of α ,

Smileyan Reductio If $\Gamma, \alpha \vdash \beta$ and $\Gamma, \alpha \vdash \beta^*$, then $\Gamma \vdash \alpha^*$.

I won't set out the rest of the rules in Rumfitt's system here. As long as we're never permitted to judge and reject the same content, we are also never permitted to judge a content and its negation (Rumfitt 2002, 310). And from the rules above (with an obvious rule for conjunction), we can deduce

Rejective Non-Contradiction $\vdash -(P \wedge \neg P)$

which mandates rejecting any contradiction. In this way, Rumfitt's rules seem to capture the point of the Law of Non-Contradiction.

If bilateralism is the right way to formally capture the Law of Non-Contradiction, which formal logics meet the test? Rumfitt's rules are classical. Nonclassical logicians will want to reject one or more of the rules in his system. Humberstone (2000, 364) and Kürbis (2016, 635) offer bilateralist intuitionist systems that reject $-\neg$ Elimination and the classical half of Smileyan Reductio – the half which allows proving a positive claim – but neither of these changes undermines the derivation of RNC. So CL and IL are in.

Bilateralism interacts interestingly with K3, LP and FDE, as suggested in the following comments by Restall (2013, 81; corner-quotes omitted):

Friends of truth-value GAPS and truth-value GLUTS both must distinguish the *assertion of a negation* (asserting $\neg P$) and *denial* (denying P). If you take there to be a truth-value *glut* at P the appropriate claim to make (when asked) is to assert $\neg P$ without thereby denying P . If you

take there to be a truth-value *gap* at P the appropriate claim to make (when asked) is to deny P without thereby asserting $\neg P$.

A thinker who allows for truth-value gaps, as in K3, will not accept $+\neg$ Introduction or $-\neg$ Elimination. If P is neither true nor false, then it is appropriate to reject P , but not appropriate to assert $\neg P$; since $\neg P$ will also be neither true nor false, it is appropriate to reject $\neg P$, but not appropriate to assert P . (For an example of the failure of $+\neg$ Introduction, we may reject a claim like ‘The King of France is bald’ without being willing to assert its negation (Rumfitt 1997; Dickie 2010).²⁰) Again, this doesn’t block the derivation of RNC.

What about someone who holds that a sentence can be both true and false, as in LP? Priest (2005, ch. 6) maintains that assertion and rejection are incompatible, which suggests some friendliness to a bilateralist framework.²¹ However, an adherent of LP would have to abandon $-\neg$ Introduction and $+\neg$ Elimination. If P is both true and false, then it is appropriate to assert P but not appropriate to reject $\neg P$. In the same circumstances, it is appropriate to assert $\neg P$ but it is not appropriate to reject P . As we would expect, abandoning $+\neg$ Elimination blocks the derivation of RNC.

Finally, someone who accepted FDE would have to drop all four rules: $+\neg$ Introduction and $-\neg$ Elimination because of truth-value gaps, and $-\neg$ Introduction and $+\neg$ Elimination because of truth-value gluts. Along with the adherent of LP, then, the follower of FDE is not committed to RNC.

If RNC is the right way to capture the Law of Non-Contradiction, then it appears that someone who followed LP or FDE would not have a general ability to express tension among their thoughts.

However, I don’t think this point is conclusive. I’ve said that adherents of nonclassical logics would have to reject certain rules in Rumfitt’s framework, but I haven’t said what would go in place of those rules. It’s possible that the substitute rules that constrain assertion, rejection and negation would give negation sufficient expressive power to underwrite a grasp of tension among thoughts. So while bilateralism seems more promising than the other formalizations we’ve considered, it does not, as presented here, offer a straightforward way of determining which formal logics contain rules suitable for thinking with. The main conclusion to draw at this

²⁰For a bilateralist framework allowing for this ‘weak rejection’, see Incurvati & Schlöder (2017).

²¹However, some say Priest should hold that assertion and rejection are compatible (Goodship 1996, 153-155; Ripley 2015). For a defence of the incompatibility, see Incurvati & Smith (2010).

point is that what it means for a deductive system to contain Non-Contradiction, in the philosophical sense at issue in my earlier arguments, is not at all straightforward. Classical and intuitionistic logic, unsurprisingly, are safe, but for all the other logics we considered, the results vary markedly depending on how the Law of Non-Contradiction is interpreted.

3 Logical capacities

3.1 Introduction

In Chapter 1, I argued that a subject's representational activity counts as thinking only if the subject manifests sensitivity to logical rules. To think any thought p – say, *that's a wild turkey over there in the woods* – you have to recognize some range of other thoughts as standing in relations of support or tension with p . For example, p supports the thought *that's a bird over there* and stands in tension with the thought *it's really a grouse*. To respond to rational relations like these, I argued, you need to follow logical rules. But I relied on an intuitive sense of what this 'following' (or, as I'll also say, 'understanding') involves. In this chapter I give an account of it. (In Chapter 2, I explained how my views can accommodate logical pluralism. In this chapter, for ease of exposition, I'll write as if there is a single correct set of logical rules.)

An account of logical rule-following has to explain how a thinker 'observes the rules of logic without thinking about them' (Ryle 1949, 47). To do this, it has to avoid two opposed problems. On the one hand, it would be implausible to suggest that thinkers have to have knowledge of logical rules, of the sort that we learn in introductory logic classes. For example, take

Modus Ponens From p and *if p , then q* , infer q .

While it might be argued that thinkers have to have tacit knowledge of Modus Ponens – a suggestion I'll return to below – it is surely not viable to suggest that thinkers have to explicitly know that Modus Ponens is a valid rule. For one thing, many thinkers have no belief to this effect, and therefore (on most accounts) no such knowledge.¹

On the other hand, Modus Ponens is not an empirical generalization about how thinkers operate. When someone reasons that it is raining, and if it's raining, the streets are wet, so the streets must be wet, it is not an accident that their reasoning

¹For another thing, for any inference-rule we can name – including rules as apparently basic as Modus Ponens – there is a philosopher somewhere who believes that the rule is not valid. For doubts about Modus Ponens, see McGee (1985) and Kolodny and MacFarlane (2010). Such philosophers are still thinking, however misguidedly.

conforms to the rule. There's some sense in which it's right to say that they reasoned this way because their reasoning was an instance of Modus Ponens. Indeed, it's hard to see how valid inference could be knowledge-transmitting – put differently, how knowledge could be closed under valid inference (Sundholm 2002; Rumfitt 2011, 352) – if this were not the case.

In one form or another, these problems have received extensive discussion in the literature on inference (see e.g. Boghossian 2003; Wedgwood 2006; Wright 2014), and I won't aim to give a comprehensive account here. The novel claim I'll make is that our sensitivity to logical rules involves *logical capacities*: inferential dispositions which display a distinctive *first-person/third-person asymmetry*. From the perspective of the thinker, certain inferences just appear compelling, without any additional reason why they are compelling. From an outside perspective, however, we can see that the inferences the thinker takes to be compelling are those which fall under a rule like Modus Ponens.

Put together with the conclusion of Chapter 1, this account of sensitivity to logical rules generates logical constraints on thinking: there are limits on how illogical we can be in our representational activity while still thinking. Precisely what those limits are the subject of the next chapter.

The rest of this chapter goes as follows. I'll start in 3.2 by saying a bit more about why sensitivity to logical rules is not to be characterized as states of explicit knowledge to the effect that those rules are valid. In 3.3, I'll suggest that logical rules must play a causal-explanatory role in our reasoning, and that this motivates a dispositional account of our sensitivity to those rules. In 3.4, however, I'll show that the simplest version of a dispositional account fails, because it leaves out the agential or first-personal character of reasoning. This leads, in 3.5, to the more sophisticated capacities account which I think we ought to accept. I offer some remarks there about whether having such capacities amounts, in any sense, to having knowledge of logic, and conclude by responding to an objection.

3.2 Cognitivism

In this section I'll briefly explain why I reject *cognitivism* about the relation between thinkers and logical rules. I won't discuss the view in detail, but will simply indicate the problems that make noncognitivist views seem more promising.

Before I begin, however, it's important to distinguish two different views that

are called ‘cognitivism’. The first view, critically discussed by Wright (2018), is that at least some thinkers, at least some of the time, have knowledge that some logical rules are valid. Such knowledge would be knowledge of certain propositions: for example, ‘Modus Ponens is valid’, or ‘If p is true and *if p then q* is true, then q is true’, or ‘it is legitimate to infer from p and *if p , then q* to q ’.

The second view, defended by Corine Besson (2012; 2019a; 2019b) is that our understanding of logical rules consists in knowledge that those logical rules are valid. In other words, what underwrites a thinker’s inference from p and *if p , then q* to q is knowledge of some proposition to the effect that Modus Ponens is valid. This proposition is ‘explicitly or consciously represented’ in the mind (2019a, 173). This is the view I will label *cognitivism*.² Importantly, according to the cognitivist, knowledge of Modus Ponens isn’t used as a premise in our reasoning – it plays a different role. It is ‘knowledge that enables reasoning’ (2019b, 61); having this knowledge ‘*gives* you a reason to infer according to MP’ (2012, 80).

Cognitivism is designed to accommodate two phenomena which crude dispositional accounts struggle with (2012, 77). First, someone who understands MP and believes p and *if p , then q* is not *bound* to infer q ; they can refrain from doing so without counting as irrational or abnormal, for example if they think that q is false. Second, it’s possible to exercise your understanding of MP without inferring according to MP – for example, you can observe that some conclusion would follow from some premises via MP without judging that conclusion.

Clearly these are requirements an account of our sensitivity to logical rules should meet. Moreover, it seems reasonable to suppose that some of us, some of the time, have knowledge of logical rules – however hard it may be to explain the etiology of such knowledge (Hale 2002; Wright 2018). But cognitivism requires the stronger claim that whenever a thinker is sensitive to a logical rule, they know a proposition which states that rule. And I think this claim is deeply implausible, even for simple rules like Modus Ponens.³

Consider a subject who reasons: it’s raining; if it’s raining, then it’s a bad day for a picnic; so it’s a bad day for a picnic. Absent some further fact which spoils

²To be clear, the two versions of ‘cognitivism’ are not unrelated: roughly speaking, if Bessonian cognitivism is true, then Wrightian cognitivism had better be true too, or else nobody ever infers in a way that reflects an understanding of the rule they conform to.

³I’m arguing here that knowing that the rule is valid is not necessary; I will not address the question of whether such knowledge would be sufficient. On the sufficiency claim, see Wright (2014), who draws connections with Carroll (1895).

the inference, this is a perfect case of Modus Ponens reasoning; provided that the subject had justified beliefs in the premises, they now have a justified belief in the conclusion. But it seems, nevertheless, to be possible that the subject does not know any proposition to the effect that Modus Ponens is valid.

On most accounts of knowledge, knowing p requires believing p . But it seems possible that this subject has no belief either way on whether Modus Ponens is valid. As Boghossian writes, this is ‘far too sophisticated a requirement’ (2001, 25). For instance, it seems possible that, if asked whether it is legitimate to infer from p and *if p , then q* to q , our subject would display no disposition to assent or dissent – they might just say they have no idea. If they have no prior exposure to formalized reasoning, or lack concepts of inference or entailment, they might not even understand the proposition (Boghossian 2001, 10). So there’s no basis for supposing that they believe that Modus Ponens is valid. Besson (2012, 79) accepts this consequence in the case of young children, pointing out that it might be reasonable to deny that young children are sensitive to MP; but the same problem applies equally to mature adults, who may be able to follow MP arguments without such propositional knowledge.

Besson (2012, 79) also appeals to the division of linguistic labour, which allows for nonexperts to count as understanding technical terms by deferring to experts: for example, I can use *elm* and *beech* without being able to tell these kinds of trees apart. She suggests that someone without much conceptual sophistication could likewise count as understanding concepts like *entailment* by deference to experts. But it seems to me that there could be a whole society in which there were no logicians at all, and so no experts to defer to. In this (sadly deficient) society, people could still be sensitive to Modus Ponens, despite lacking even a deferential grasp of concepts of implication or entailment.

A different line of response is that, ordinarily, we are said to believe not only the propositions that we are disposed to assent to, but also the consequences – at least the direct consequences – of those propositions. But it’s not clear that invoking this closure condition helps here. It would help only if the validity of Modus Ponens was a consequence of the subject’s prior beliefs: but the subject may have no prior beliefs about logic whatsoever, and therefore no prior beliefs that entail the validity of Modus Ponens.

Perhaps it could be argued that the validity of Modus Ponens is a logical truth, and therefore a consequence of any set of beliefs at all. But this is a desperate move. If the only fact in virtue of which the subject counts as believing p is that

p is logically true, the subject's belief in p can hardly be cited to explain anything further, such as the validity of the subject's reasoning. In any event, holding that a subject believes in Modus Ponens on this basis would commit us to accepting that the subject also believes in every other logical truth on the same basis – not an attractive position. Moreover, if the subject counts as *knowing* the validity of Modus Ponens on this basis, then they should also count as knowing every other logical truth on the same basis. In other words, the subject must be logically omniscient.

The same is true if we require a thinker to know only the weaker claim that the particular premises they are inferring from entail the particular conclusion that they are inferring to. Again, it seems that a subject could make an inference without having the requisite beliefs, or even the concepts needed to form such a belief (Boghossian 2014, 7-8). In response to this point, Zalabardo (2011, 126) argues that '[a] subject's inclination to infer q from p should normally be taken as sufficient grounds for ascribing to him a belief in the existence of at least a probabilistic link between the truth of p and the truth of q ', even if the subject isn't able to consciously entertain the proposition that p entails q . But while it's plausible that an inference requires such an inclination to infer, there's no good reason to redescribe the inclination as a 'belief', especially in subjects who are unable to make the inclination explicit in a propositional form. Unlike an ordinary belief, which is inferentially integrated with other beliefs, evidence and action (Evans 1982, 336), the inclination Zalabardo appeals to is not manifest in any role other than the particular inference it is about. If an inclination to infer suffices for the truth of cognitivism, then cognitivism is trivial, and we are better off investigating the nature of the inclination to infer.

So it seems doubtful that anyone who infers validly in accordance with Modus Ponens has a belief that Modus Ponens is valid, or even a belief that a particular instance of it is valid – let alone knowledge of one of these things. Our sensitivity to the rule, then, cannot consist in such knowledge.

I'm not suggesting that this objection is conclusive. There's more the cognitivist might say – either to support the attribution of logical knowledge even in cases where the thinker disclaims it or is unable to articulate it, or to undermine the sense that a thinker who overtly rejects a logical rule can nonetheless follow that rule in inference. But I think this is enough to show that noncognitivist approaches are worth investigating. At the end of the chapter, I'll suggest that there is a sense in which we know logical rules, but not the sense appealed to by the cognitivist.

3.3 The causal-explanatory role of logic

In this section, I argue that logic has a causal-explanatory role in reasoning, and use this to motivate a simple dispositional account of our inferential capacities.

By ‘the causal-explanatory role of logic’, I mean the following. Judging p , *if* p , *then* q and then q may not amount to an inference. An inference requires that, in some sense, you judged the conclusion because you judged the premises. Implicit in this requirement is a causal condition: there must be a causal relation between judging the premises and judging the conclusion. What’s more, this relation must be of the right sort: judging the premises must cause you to judge the conclusion because the premises entail the conclusion. As Wedgwood (2006) puts it, the beliefs in the premises must cause the belief in the conclusion in virtue of rationalizing it. Or compare what Peacocke (2004, 175) calls the ‘fixed-point principle’: ‘It is a condition for a judgment to be rational that it be made in part because it is rational’. In this section, I’ll work my way towards the causal-explanatory role of logic by considering some accounts that reject it.

3.3.1 Dogramaci’s associationism and the sensitivity condition

It seems obvious that, in inference, you judge the conclusion because you judged the premises; but it’s not so obvious that this casual relation must hold ‘in virtue of rationalizing’ the conclusion. One way of rejecting this further condition is to maintain that inference requires only a certain causal relation between judging the premises and judging the conclusion, but that the logical entailment plays no role in this causal relation. Sinan Dogramaci has proposed an associationist account of deduction which has this shape. He seeks ‘to unite the basis and conclusion of an inference not by appealing to a consequence relation, but something psychological, a mental state’ (2013, 392).

On Dogramaci’s account, you are in a position to infer when you stand in a certain ‘unmediated psychological relation’ to the premise-judgments and the (potential) conclusion-judgment. It is a three-place relation between the subject, the premise-judgments and the conclusion-judgment: the subject is tempted by the premise-judgments to judge the conclusion. Dogramaci calls this relation a ‘conditional intuition’. Inference, then, works like this: you judge p and *if* p , *then* q ; these judgments causally generate an intuition – a temptation to judge q . Succumbing to this temptation, you judge q . The existence of the conditional intuition is necessary

and sufficient for inference: the fact that the premises entail the conclusion has no explanatory role.

As it stands, this account faces a simple problem. Since having a conditional intuition doesn't depend in any way on the validity of the inference it tempts you to, there can be all sorts of defective intuitions. Suppose that you are tempted to deny the consequent: whenever you judge *if p, then q* and *not-p*, you are tempted to judge *not-q*. If you succumb to that temptation, then all the conditions for inference are in place. But such a move doesn't transmit justification, even if *not-q* turns out to be true.

We might try to improve on Dogramaci's position by holding that inference requires, in addition to (i) a causal connection between the premise-judgments and the conclusion-judgment, (ii) an entailment between the premises and the conclusion. This is consistent with denying that the entailment plays a role in the causal connection: the two conditions are independent.

I don't think this will work. As Wedgwood has shown, this sort of view is subject to problems about deviant causal chains: it allows for the possibility that 'it is just a fluke that these antecedent mental states both rationalize and cause the formation of the new belief' (2006, 670). For example, suppose that any time you judge a conditional, you are tempted to judge the consequent of that conditional (i.e. judging *if p, then q* tempts you to judge *q*). Most of the time, such intuitions will lead you astray, but once in a while you might also – independently – have judged the antecedent of the conditional, *p*. In such cases, according to the revised account, your inference will transmit justification. But since you would've had the conditional intuition whether or not you judged *p*, the intuition can do no justificatory work. The idea underlying Wedgwood's counterexample is similar to the sensitivity condition in epistemology (Nozick 1981, 179): you know something only if, had it been false, you would not have believed it.⁴ Similarly, your inference is justified only if, had the premises not entailed the conclusion, you would not have been tempted to draw it.

We could simply add this sensitivity condition to the current account. Then an inference would be justified if and only if (i) the subject is caused by judging the premises to judge the conclusion, (ii) the premises entail the conclusion and (iii) if the premises had not entailed the conclusion, the subject would not have been caused by judging the premises to judge the conclusion. This account is not wrong,

⁴For the closely related safety principle, which holds that you know something only if you could not easily have been wrong in a similar case, see Williamson (2000, 147).

but it's also not illuminating. In adding the sensitivity condition, we have in effect admitted that the entailment between the premises and conclusion has some role in shaping the causal relation between the subject's acts of judging. But, as (iii) is stated in counterfactual terms, the account fails to explain how this sensitivity comes about. This is precisely what I want to explain.

3.3.2 The simple dispositional account

We can unify conditions (i), (ii) and (iii) at a deeper level by accepting the causal-explanatory role of logic. An inference is justified only if the fact that the premises entail the conclusion plays an appropriate role in explaining why the subject was caused by judging the premises to judge the conclusion. As Wedgwood (2006, 670) writes, 'These antecedent states must cause one to form that belief or intention precisely in virtue of their rationalizing one's forming that belief or intention.'

This point has been recognized elsewhere. For example, in a discussion of belief-desire explanation of action, Jerry Fodor notes that 'causal relations among propositional attitudes somehow typically contrive to respect their relations of content' (Fodor 1987, 12). This, he says, is why, when someone reasons well, a description of the causal history of their thinking is also a description of a valid argument. The problem is to explain how exactly this could hold. As Fodor asks (1987, 14): 'how could the mind be so constructed that [the generalizations that subsume belief/desire causation] are true of it? What sort of mechanism could have states that are both semantically and causally connected, and such that the causal connections respect the semantic ones?'⁵

As Fodor recognized, it's difficult to devise such a mechanism when it comes to all the 'relations of content' which could underwrite good reasoning. There are just too many such relations for any set of dispositions to plausibly capture. Luckily, we're only aiming to give an account of logical reasoning, and for this a dispositional story seems entirely apposite. The story would go as follows. Recall the inference rule

Modus Ponens From p and *if p , then q* , infer q .

⁵Note that this is not the general problem of how something that is not an event can have a causal role. I take it that, even if only events can literally be causes, things other than events can be part of a causal explanation. For example, the properties of my soup and the relation between the soup, the pot and the stove can both be part of the causal explanation of the soup's boiling over, although neither a property nor a relation is an event.

The capacity to infer according to Modus Ponens consists in a disposition which is activated when a subject has two beliefs, of the form p and *if p , then q* . The result of the disposition's being activated is that the subject then judges that q . Call this 'the simple dispositional account'.

Two features of the simple dispositional account deserve emphasis. First, the disposition is a causal intermediary between the event of the subject's judging p and *if p , then q* and their judging q . This allows us to say that the subject judged q because they judged p and *if p , then q* . Second, the disposition is defined such that it operates properly only when the subject is in a position to draw an inference validated by Modus Ponens. So the fact that the premises entail the conclusion is part of the explanation for why the subject judged the conclusion. This allows us to say that the subject was caused by the premises to judge the conclusion in virtue of the fact that the premises entail the conclusion. The simple dispositional account perfectly captures the causal-explanatory role of logic.

Of course, a fuller development of the account would need to further qualify the description of the disposition: it's not as if, any time two of my beliefs entail a third proposition via Modus Ponens, I draw that inference. There must be further activation-conditions: for example, it seems plausible that the subject has to consider the premises together in order to be disposed to draw the inference (Besson 2009; 2012; Murzi & Steinberger 2013). But I won't spend any more time on these details, as the simple dispositional account faces a more serious problem.

3.4 The first-personal character of inference

Earlier, I quoted Peacocke's fixed-point principle that a rational judgment must be made because it is rational. Peacocke adds that this means 'rational from the thinker's own point of view' (2004, 176). In this section, I will argue that the simple dispositional account leaves out the *agential* or *first-personal* character of inference. On the simple dispositional account, inference happens mechanically when an entailment is available. In reality, however, inferring is not just something that happens, but something we do (Boghossian 2014, 16). (I should emphasize, however, that 'agency' here is not meant to evoke explicit choosing; rather, I mean activity which is not automatic and for which the agent is responsible. Having agency, in this sense, is closely related to having control.)

To motivate this requirement, consider a case described by Bilgrami (1998):

the case of Oblomov, the wholly passive subject.⁶ When Oblomov ‘thinks’, this thinking is not something he does; rather, he is ‘assailed’ by thoughts. In other words, Oblomov is subjectively aware of thoughts, but experiences them as simply happening to him rather than as being under his rational control. In some cases, Oblomov may undergo a series of thoughts which stand in logical relations: for example, he is assailed by the thought p , and then by *if p , then q* , and finally by q . But he does not experience the first two thoughts as providing a reason for him to have the third. Now, let us suppose, further, that Oblomov has the disposition proposed at the end of the previous section. Given this supposition, the occurrence of q in Oblomov’s mind is caused by the occurrence of p and *if p , then q* , in virtue of the fact that p and *if p , then q* entail q . In other words, Oblomov’s series of thoughts here satisfies the simple dispositional account. It still seems to me that Oblomov has not drawn an inference.

There are different ways of saying what is missing in Oblomov’s case. We could say that his sequence of mental events lacks what Rödl (2018, 91ff.) calls the ‘first-personal necessity’ of inference. While the first two thoughts necessitate the third, there is nothing first-personal about this necessitation: Oblomov does not experience the first two thoughts as necessitating *him* to judge the third. Similarly, Korsgaard (2009, 69) writes that when you draw an inference,

it’s an act of self-determination, in the sense that the activity of your own mind is part of what produces the belief in you. Suppose you believe two premises, and a certain conclusion follows. You won’t automatically believe that conclusion, because you might not notice the connection between them. But if you do notice the connection, and put the premises together in the way suggested by the connection, then you *do* something: you *draw* the conclusion. In drawing the conclusion—or, as we say, in making up your mind, in *constituting* your mind—you determine yourself to believe it.

While Oblomov is disposed to conform to Modus Ponens, there’s no sense in which he makes up his mind in coming to the conclusion.

Burge expresses the same point in the following remarks (1998b, 251):

[I]n reasoning, no thinker can be a mere observer of reasons and their effects on reasoning. For having reason requires at some point having some tendency to be affected by reason’s power in motivating reasoning.

⁶Bilgrami’s Oblomov is ‘an extreme version of’ the main character from Ivan Goncharov’s 1859 novel *Oblomov*. For further discussion of the case, see Rattan (2002, 149).

Understanding what a reason is, is partly understanding its motive force ... Thinking is necessarily associated with reasoning—thinking guided by reasons—and reasoning cannot in general be a mere ‘going on’. In making inferences, a being is *ipso facto* an agent. ... In recognizing the effect of reasons on one’s judgements and inferences, one cannot reasonably think of oneself as powerless.

Burge’s discussion suggests that Oblomov doesn’t even count as understanding the logical relations among his thoughts, because a condition for understanding these relations is that he experience judging p and *if p , then q* as motivating him to judge q .

A final characterization of what is missing here comes from Moran (2001, ch. 2), who emphasizes that our attitude towards our beliefs is normally not spectatorial, but deliberative: the job is not to find out what we already believe, but to decide what to believe. (This is why we can ordinarily determine whether we believe p simply by asking ourselves whether p .) Oblomov could, at best, relate to his mental processes in a spectatorial way – as revealing to him what he already believes – but not as a process of deliberation, of deciding what to believe.

The point all these discussions bring out is that inference is not only something that happens, even something that happens within our awareness, but something we do. Oblomov’s case shows that the simple dispositional account could be satisfied while this condition was not met. So that account must be rejected.

3.5 Logical capacities and tacit knowledge

In this section I will offer a more sophisticated account of our relation to logical rules which captures both the causal-explanatory role of logic and the first-personal character of inference. As with the simple dispositional account, we will suppose that subjects have dispositions which mirror the logical rules that validate their inferences. This time, however, we will suppose that the dispositions have a distinctive sort of structure – the structure of what I will call a *capacity*.

The notion of a capacity, as I am using it here, has a built-in asymmetry between first-personal and third-personal aspects, such that a description of the capacity from only one of these aspects will be incomplete. From a third-personal point of view, the capacity for inferring according to Modus Ponens is defined by that rule, which states what is done by a subject exercising the capacity. If we observe the inferences drawn by the subject who has the capacity, we can see that they tend

to fall under the pattern specified in Modus Ponens.⁷ But the capacity also has an ineliminable first-personal aspect. And Modus Ponens does not express what it is like for a subject to exercise the capacity. For it is stated at a level of generality that need not be available to the subject exercising the capacity. The first-personal description on its own, however, will fail to show the capacity's relation to the rule – as when we say that, from the subject's point of view, it is just a brute fact that *this* inference is compelling.

This feature of our relation to valid inferences is described nicely by Zalabardo (2011, 130-131):

All of us, even the logically illiterate, have the ability to recognize valid arguments, to a greater or lesser extent. Some of us are very good at it. At least in the simplest cases, and outside our logic class, judgments of validity can be perfectly immediate. ... [I]n the multitude of everyday cases in which we recognize valid arguments without formulating the underlying rules of inference, sometimes even without being capable of formulating them, it is natural to say that our ability to recognize valid arguments gives us knowledge, and that this knowledge is non-inferential. And this will be so even if the sub-personal cognitive devices that sustain this ability operate by recognizing certain universal inference patterns.

I don't quite agree with Zalabardo's framing of the point, though: we may not be able to formulate the underlying rules, but that need not make our grasp of them sub-personal. (I make some related comments in the next section.) I prefer to frame the point in terms of a first/third-person distinction: first-personally, we can immediately take an inference to be valid, while third-personally, we may be recognizing it as an instance of a rule.

Let me say a bit more about the first-personal dimension of the capacity. The subject exercising the capacity judges p and *if p , then q* and then moves from these to q . To describe what this is like first-personally, we might say, in Peacocke's phrase, that once the subject has judged the premises, they find the conclusion 'primitively compelling' (1987, 155). That is, they find it rationally compelling to judge the conclusion given their belief in the premises, but primitively – not because of some further belief, such as a belief in Modus Ponens or a belief that the premises entail the conclusion.⁸ A subject who has the capacity will find compelling the transitions that

⁷I say 'tend' because a capacity may sometimes misfire. Just how much a capacity can misfire before it fails to count as a capacity at all is a question I consider in 4.2.1.

⁸Isn't it possible, though, for an intuitionist to understand classical negation without finding

are caught by the associated rule, but not *as* transitions caught by the rule. Rather, there need not be anything further available to the subject than that this inference is an attractive one. In other words, the subject with the capacity is tempted to draw particular inferences which evidently fall under the rule; they need not be tempted to accept a general representation of the rule.

We may also want to allow for capacities where a subject finds a particular content, or set of contents, ‘primitively unacceptable’. For example, this seems like a plausible description of our relation to the Law of Non-Contradiction: when confronted with a combination of beliefs which are explicitly contradictory, we find it rationally unacceptable to hold on to all of the contradictory beliefs, but primitively – not because of some further belief. Murzi (2020, 412) describes the process as follows: ‘upon deriving both A and $\neg A$, a rational agent *stops* her reasoning and examines instead which of the assumptions on which A and $\neg A$ depend must be given up.’ But whether we need to recognize a separate kind of capacity for this kind of case, or whether primitively compelling inferences suffice, depends on how exactly the Law of Non-Contradiction should be represented – a topic I discussed in 2.4, but without reaching a definitive conclusion. In what follows, then, while I’ll refer to ‘primitively compelling’ inferences for readability, this should be taken to leave open the possibility of primitive unacceptability.

Along similar lines, Harman (1986, 18) proposes that we have fundamental dispositions ‘to treat some propositions as *immediately implying* others and some as *immediately inconsistent* with each other’. For present purposes, I want to adopt Harman’s observation that (in my terminology) finding an inference from p to q compelling need not require that the compulsion be absolute or overriding: for one thing, ‘[the subject]’s general disposition may be overridden by other considerations in a particular case, for example, if [q] is absurd’ (19). Similarly, a disposition to avoid believing contradictories ‘may be overridden in a particular case, as when [the subject] is disposed to believe the premises of the liar paradox’. And finally, one may not find compelling a transition which in fact falls under the rule if the particular thoughts are too ‘long or complex or otherwise distracting’ – or, I think we can add, if one is simply tired or not paying attention.⁹

instances of specifically classical reasoning at all compelling? I think this is possible; it’s what Moore (2003b, 48) calls a ‘disengaged’ grasp of a concept. But for logical concepts to constitute our responsiveness to rational relations among thoughts, one must grasp them in an ‘engaged’ way.

⁹For more interesting cases in which such compulsions are overridden, see Moore (2003b, 49). Moore argues that possessing the concept of a swear word requires more than ‘knowing, in some

Field (2009, 260) adds that some kinds of inference are only obvious once you consider both the premises *and the conclusion*; if the conclusion isn't brought to your attention, then merely believing the premises doesn't tempt you to draw the conclusion. The clearest example is Universal Instantiation: believing *Every politician exaggerates* doesn't tempt you to believe *Jean Chrétien exaggerates* until Jean Chrétien is brought to your attention. But it may be that other inference-rules display the same phenomenon to a lesser degree.

We can characterize the first-personal dimension of a capacity in terms of *reasons*: if p immediately implies q , then a subject who believes p generally has a reason to believe q as long as they continue to believe p ; similarly, if p is immediately inconsistent with r , then a subject generally has a reason not to believe both. This helps to bring out the sense in which an exercise of a logical capacity is rational from the subject's point of view. But this way of characterizing things needs to be qualified. The 'reason' that the subject has merely in virtue of having a logical capacity is not a justifying reason, but an explanatory one: it is a motivational state which could explain the subject's action (Parfit 1997; Alvarez 2017). A subject might, for all I've said here, have a logical capacity to follow the rule for *tonk*; in this case, the subject would find instances of the *tonk* rules primitively compelling, and might sometimes draw these inferences, but would not be justified. So it's not built in to the nature of a capacity that it generate normative reasons, but it is built in that it generate explanatory reasons. (Of course, where the rule is a good one, the subject may *also* have a normative reason to draw the inference, but this doesn't follow merely from having a capacity to follow a rule.)

So primitive compellingness and unacceptability pick out a psychological phenomenon, not a normative one. But, while compellingness and unacceptability do normally manifest in subjective feelings that one should judge (or not judge), they aren't identical with such feelings. One might have a subjective feeling that one should judge (or not judge) that arises from misunderstanding of the thoughts in question, such that the feeling disappears when one comes to better understand the thought. Or one might *not* have a subjective feeling when one misunderstands a thought, but come to have the feeling when one understands it better. Frege draws a similar distinction between whether a thought is really self-evident or merely seems

aloof, anthropological way, that the use of these words in certain contexts will shock certain people. To possess the concept of a swear word one must have *some* tendency to share that shock, even if one habitually and enthusiastically flouts it, or even if one habitually and nonchalantly flouts it.'

so, and allows that a self-evident thought that is not obvious to us can become obvious as we come to understand it (Burge 1998a, 342-343; Jeshion 2001, 970-971).¹⁰

Failure to recognize the first-personal/third-personal asymmetry in our inferential capacities is, in my view, the root of much puzzlement about the nature of those capacities. Let me give two examples. First, Markos Valaris (2017) has objected to rule-following accounts of reasoning on the following basis. Reasoning is a way of ‘making up our minds’ about how things are in the world; in reasoning, our attention is not on our thoughts but on the world. But logical rules are about relations among thoughts, so if reasoning involved following logical rules, then in reasoning, our attention would be on our thoughts, not on the world. So reasoning does not involve following logical rules. In my view, the assumption that reasoning is either about the world or about relations among our thoughts fails to allow for the distinction between first-person and third-person descriptions of reasoning. First-personally, a subject’s experience of reasoning may simply involve attention to what is the case. This is consistent with the subject’s activity still being an activity of rule-following – a fact which holds third-personally, where it turns out that the subject draws inferences which are validated by logical rules.

Second, Besson has recently defended the idea that we directly apprehend the validity of particular inferences, without going by way of any general principle (2019a, 189): ‘Your reasoning need not go through the recognition of a general logical principle, it can be direct.’ She points to ‘a long tradition of taking rational insight to apply directly to particular propositions/arguments and taking the formal principles to be secondary in the order of apprehension and justification of such propositions/arguments’ (2019a, 188), citing Russell, Cook Wilson and Bonjour. She denies that insight into a particular inference requires knowledge of a logical rule, suggesting instead that our knowledge of logical rules may involve induction from particular instances of valid inference.¹¹

I think it’s quite right that, first-personally, we directly apprehend the validity of particular inferences. But if grasp of a general rule plays no role in this process, then it’s puzzling how we manage to apprehend the right inferences, especially when it comes to an inference we’ve never considered before. What could give us this insight,

¹⁰I’m not suggesting that the notion of compellingness entirely overlaps with Frege’s notion of self-evidence. Frege’s notion is embedded in the rationalist conception of a ‘natural order of truths’ (Frege 1884, §17; Leibniz 1765, 412), which sits awkwardly with the holism of Chapter 1.

¹¹Besson has confirmed (p.c.) that this modifies the view in her earlier papers, although it’s still intended as a form of cognitivism.

if not the rule? The first-person/third-person asymmetry in our logical capacities shows that we can maintain both that thinkers directly apprehend the validity of particular inferences and that they apprehend them as instances of general rules.

So we should accept that understanding logical rules involves logical capacities in this sense. This fact has consequences, which I explore in the next chapter, for the extent to which someone can violate logical rules and still count as thinking.

3.5.1 Do thinkers have logical knowledge?

Finally, I want to offer some comments on whether having logical capacities amounts to knowing logical rules in any sense. Clearly, having a logical capacity to infer according to Modus Ponens does not amount to having explicit knowledge that Modus Ponens is valid. But could it amount to knowledge *of* Modus Ponens in some other sense? There is some reason to think that it could. I described logical capacities in terms of a first-person/third-person asymmetry, but the same structure could also be described in terms of a distinction between a practical ability, which issues in a sense of particular actions as appropriate or inappropriate, and the theoretical representation of that practical ability, which displays it as grasp of a set of general principles. And this structure is what many philosophers take to be characteristic of *tacit knowledge*.

The notion of tacit knowledge is familiar as a characterization of the linguistic competence of ordinary speakers. Suppose that English grammar can be described by a systematic theory of meaning (Dummett 1975). While learning to speak English does not involve explicitly learning this theory, we might still hold that the English speaker has tacit knowledge of the theory, because the explanation for the speaker's competence parallels the structure of the theory: as Davies (1989, 132) writes, 'tacit knowledge of a particular systematic theory is constituted by a causal-explanatory structure in the speaker which mirrors the derivational structure in the theory'.¹² In virtue of this causal-explanatory structure, the speaker is sensitive to the information specified by the rules of the grammar. We can therefore say that the grammar is psychologically real for the subject, as Peacocke writes, since 'for a rule of grammar to be psychologically real for a given subject is for it to specify the information drawn upon by the relevant mechanisms or algorithms in that subject' (1989, 114).

¹²See also Davies (1981, chs. III and IV), Evans (1981), Fricker (1982), Peacocke (1986) and, for objections, Quine (1970a) and Wright (1987, ch. 6). The connection between inference and these debates is picked up in Engel (2005, 34).

So I propose, tentatively, that logical capacities are states of tacit knowledge. Someone who has a logical capacity is sensitive to the features of thoughts specified by the relevant logical rule: they find transitions that fall under the rule primitively compelling. In virtue of having these capacities, the causal-explanatory structure underlying the thinker's activity mirrors the logical rules stated in a description of the capacity. As with other kinds of tacit knowledge, a statement of the associated rules will go beyond what it is like for the thinker exercising the capacity. This is not to say that it doesn't matter how we specify what is known; we should try to specify it in a way that brings out the features which cause particular transitions to appear primitively compelling. Still, it may be that there are multiple, equally correct ways of specifying the content of a thinker's tacit knowledge of logic. For example, there might be no basis for choosing between various proof-theoretically equivalent sets of rules.

3.5.2 Logical capacities and separability

According to the account I've just given, a subject is sensitive to a logical rule if, and only if, the subject has a logical capacity to make the transitions specified by the rule. In this section, I'll briefly consider a formally motivated concern that this position might raise: that the rules picked out by logical capacities have to be separable, and therefore can't be classical.

The account I've given relied on a conception of logic as a set of rules. Any such account has to explain the fact that not every set of inference-rules determines a genuine logical connective. For example, consider the rules defining Prior's (1960) connective *tonk*:

Tonk Introduction From p , infer $p \text{ tonk } q$.

Tonk Elimination From $p \text{ tonk } q$, infer q .

These rules are trivializing: if added to a language in which something can be proved, they allow everything to be proved. In response to this kind of problem, inferentialists have proposed additional constraints on which rules are legitimate. Two such constraints are harmony and separability.

Harmony, a requirement proposed by Dummett (1991, 219) and deriving from Gentzen (1969, 80), is meant to capture the idea that deductive inference serves to transmit evidence from thought to thought, so the *grounds* which license you to

infer a thought of some form should be in harmony with the *consequences* which you are licensed to infer from a thought of that form.¹³ The rules for *tonk* are clearly disharmonious, but there are problems in giving a harmonious formalization of classical logic (Dickie 2010). I won't say any more about harmony here, however; even if it is a defensible constraint on inferential rules, nothing in my own account creates additional pressure to satisfy it, or sheds additional light on it.

The situation is different with *separability*, which is meant to capture an '*atomistic* conception of our understanding of logical expressions', where 'in principle a speaker could understand e.g. \wedge without understanding \exists , \rightarrow without understanding \neg , and so forth' (Murzi 2020, 395). As Tennant writes, one should

be able to master various fragments of the language in isolation, or one at a time. It should not matter in what order one learns (acquires grasp of) the logical operators. It should not matter if indeed some operators are not yet within one's grasp. All that matters is that one's grasp of any operator should be total simply on the basis of schematic rules governing inferences involving it. (2002, 319)

Separability is a general constraint on a set of logical rules. It requires that 'every provable sentence or rule in the system has a proof that only involves either structural rules or rules for the logical operators that figure in that sentence or rule' (Murzi 2020, 395).¹⁴ My talk of logical capacities may seem to lend additional support to separability, because it appears that the capacity involved in understanding negation is distinct from the capacity involved in understanding, say, the conditional, so it should be possible to have one without the other.

The reason this might be problematic is that there are well-known difficulties in giving a separable formalization of classical logic. For example, consider the following classical theorem:

Peirce's Law $((P \rightarrow Q) \rightarrow P) \rightarrow P$

Peirce's Law contains no connective other than \rightarrow , but it cannot be derived from the standard rules for \rightarrow alone; it requires using the rules for \neg too. This is a failure of separability. Such failures have been used to argue for accepting nonclassical logics (Dummett 1991; Tennant 2002, 319ff.). If logical capacities have to specify separable rules, does this mean that the rules that our capacities specify cannot be classical?

¹³For formulations of harmony, see Steinberger (2011a) and Murzi & Steinberger (2017). For objections to harmony, see Rumfitt (2017).

¹⁴As Murzi notes, it is equivalent to the conservativeness requirement from Belnap (1962).

That would be too quick, for two reasons. First, it is possible to give a (harmonious and) separable formalization of classical logic, although always with some nonstandard features. One option is to permit *multiple conclusions*, such that the most general form of an argument is that if all of the premises are true, then at least one of the conclusions is true (Restall 2005). But it seems implausible that our capacities really specify such rules: we don't tend to make multiple-conclusion arguments in ordinary reasoning (Steinberger 2011b).

A more promising option is *bilateralism*, which treats *rejection* as a primitive kind of force on par with assertoric force, such that rejecting p is not reducible to asserting *not-p* (2.4; Rumfitt 2000). Allowing for inference rules to constrain patterns of assertion (judgment) and rejection makes it possible to give a harmonious and separable formalization of classical logic. This would suggest that we have logical capacities not only to judge a thought on the basis of judging others, but, more broadly, to judge or reject a thought on the basis of judging or rejecting others. Developing my account in a bilateralist way would require recognizing primitive unacceptability as the first-personal aspect of following a rule that requires rejecting a conclusion, just as primitive compellingness is the first-personal aspect of following a rule that requires accepting a conclusion.

A third option comes from Murzi (2020), who offers a different formalization of classical logic which uses single conclusions and only assertoric force. Murzi makes two nonstandard moves. Absurdity, \perp , is treated as a 'logical punctuation sign', rather than a sign for a proposition which could occur in proofs, though we are never entitled to assert it (cf. Tennant 1999). In other words, it represents the way that, 'upon deriving both A and $\neg A$, a rational agent stops her reasoning and examines instead which of the assumptions on which A and $\neg A$ depend must be given up' (Murzi 2020, 412). And just as in standard natural deduction systems we can assume some proposition and discharge the assumption in the course of a proof, in Murzi's system we are allowed to assume a *rule* and discharge that assumption; similarly, we can arrive at a rule as the conclusion of a proof. The activities of introducing and eliminating rules are governed by 'higher-level rules' – rules for inferring to and from rules.

In fact, these higher-level rules are not as exotic as they might seem. Geach (1976, 87) refers to the Stoic distinction between *schemata*, which go from propositions as premises to propositions as conclusions, and *themata*, which go from arguments to arguments, or, as we might say, from rules to rules. He takes our ordinary

practice of reasoning to include both; an example of an ordinary thema would be ‘a plaiting together of two arguments into a chain, in which the ultimate conclusion is taken to follow from the original premises’ (Geach 1976, 88). Higher-level rules are a generalization of schemata and themata, since they can relate both rules and propositions as premises and conclusions. To develop my account in this way, we would have to recognize capacities that take rules as inputs and outputs. If Geach is right, however, we already have independent reason to recognize such capacities as constituting our grasp of the structural rules which govern how arguments can be put together.¹⁵

The first point, then, is that there are separable formalizations of classical logic, so if logical capacities have to specify separable rules, this doesn’t mean that the rules that our capacities specify cannot be classical. Second, however, I doubt that logical capacities have to specify separable rules, or even that my account lends any additional support to separability as a constraint on a set of logical rules.

Separability requires that we could fully understand some connective \otimes solely in virtue of our sensitivity to the rules for \otimes , without understanding any other connectives. But the mere fact that logical capacities can be separately identified (by the rules they pick out) does not entail that, say, the capacities involving \rightarrow are enough on their own for a full understanding of \rightarrow . It might just be the case that, to understand everything about \rightarrow , you also need to understand \neg .

Brandom (1994, 127) rejects conservativeness as a general constraint on introducing concepts, observing that nonconservativeness

just shows that [the concept added] has a substantive content, in that it implicitly involves a material inference that is not already implicit in the contents of other concepts being employed. ... Conceptual progress in science often consists in introducing just such novel contents.

He endorses separability, however, as a constraint on logical concepts, because he thinks the point of logical concepts is to make explicit the material inferences we are already committed to: if the logical rules are not conservative, ‘the introduction of the new [logical] vocabulary licenses new material inferences, and so alters the contents associated with the old vocabulary’ (2000, 68). But if logical concepts are, *pace* Brandom, necessary for thinking, then it shouldn’t be surprising if they form a nonconservative extension of a conceptual scheme without them, as they usher in a

¹⁵There might be reason to restrict our ability to plait together arguments when their premises contradict each other: see Tennant (1978, 183-4). But *some* structural rules will be required.

new kind of representational activity.

More generally, the fundamental role that logical connectives play in our grasp of thoughts should make it unsurprising if they display a local holism, where understanding one connective requires understanding others. Priest (2006, 92) writes that ‘[t]he non-conservative nature of the extension may be attributed to the fact that the old set of rules was simply incomplete’. A set of rules with a conditional but no negation, or vice versa, may simply be incomplete.¹⁶ This kind of holism also occurs elsewhere in our basic conceptual scheme: for example, although the concepts *object* and *property* are separately identifiable, it would be impossible to fully understand one without the other; it seems unlikely that we could derive all the conceptual truths about objects without invoking the concept of property, or vice versa.

None of this is meant to show that separability must be rejected. But I think it is enough to show that accepting logical capacities doesn’t provide further support for separability, and that even if we do accept separability, this doesn’t rule out classical logic.

¹⁶There might be independent reasons for accepting this result. For example, it has been argued in connection with Kripke’s *adoption problem* that we cannot understand any rule of inference without understanding Universal Instantiation and Modus Ponens, because any rule of inference is both general and of conditional form (Padró 2015; Finn 2019b). But UI and MP are themselves rules of inference. So UI and MP form a local holism, where understanding either requires understanding the other.

4 On the limits of illogical thinking

4.1 Logic as constitutive of thinking

In Chapter 1, I argued that a subject's representational activity counts as thinking only if it manifests sensitivity to logical rules. In Chapter 3, I argued that this sensitivity consists in the exercise of logical capacities. So thinking requires exercising logical capacities. Conversely, if you aren't exercising logical capacities, you aren't thinking. In this sense, logic is *constitutive* of thinking.¹ In this chapter, I'll discuss the limits on illogical thought that follow from this view.

The claim that logic is constitutive of thinking, in the sense I've argued for, can be found in several different places in recent philosophy (generally of Kantian extraction). AW Moore (2003b, 92) expresses it as follows in his commentary on Kant's moral and religious philosophy:

[O]ne does not count as *thinking* unless one has an inclination of this sort, an inclination to yield to the demands that rationality places on thought. This is, in part, what thinking is. Thinking essentially involves the exercise of concepts. And the exercise of concepts is not possible without due respect for the demands of rationality, since these determine what counts as keeping faith with concepts and thus what counts as exercising them at all. ... Exceptional episodes of irrationality in one's thinking must be precisely that: exceptional. They must be episodes in which some isolated failure on one's part to make sense fits into a larger pattern of sense that one makes.

Christine Korsgaard claims, similarly, that '[t]he laws of logic govern our thoughts because if we don't follow them we just aren't thinking' (2009, 32), although she is clear that this doesn't mean such laws are 'impossible to violate' (30). And in *Making it Explicit* (1994, 636), Robert Brandom argues that 'To be in the game at

¹The sense of 'constitutive' here is the same as that of Korsgaard (2008; 2009). It is close to the sense of 'constitutive' used by MacFarlane (2002) and Leech (2015), except that they take it to require full, rather than partial, conformity to logical rules. It is distinct from what MacFarlane (2002) and Leech (2015) call 'constitutive normativity', which does not require any conformity. It is also distinct from the usage of 'constitutive' in contemporary metaphysics to refer to the relation between a material object and its matter (e.g. a statue and its clay).

all, one must make enough of the right moves – but how much is enough is quite flexible.’

All of these philosophers hold that thinking requires *some* degree of logicity, but none of them take a definite position on just how much and what kind of illogicality is possible before we fail to think. And without clarity on this issue, we don’t know exactly what it means to say that logic is constitutive of thinking. In this chapter, then, I explain what the limits on illogical thinking are.

I’m going to argue that there are two different limits on illogical thinking. First, thinkers must *have the capacities* to follow logical rules. This requires that they *tend* to find instances of logical rules compelling. Second, thinkers must *exercise* those capacities. This requires that they not think in *obviously* illogical ways. Clarifying the relevant notions of *tendency* and *obviousness* is the job here.

4.2 The first limit on illogicality

In this section I argue for the first limit on illogical thinking:

First Limit Thinkers must tend to find instances of logical rules primitively compelling or primitively unacceptable.

Without this tendency, a subject doesn’t count as having logical capacities at all. And without having logical capacities, the subject is not thinking. I’ll clarify what this limit means by considering a worry: that it is committed to drawing an arbitrary line between thinking and nonthinking.

This limit on illogicality marks a failure of competence, not of performance. I argued in the last chapter that someone who has a logical capacity will find certain judgments or inferences primitively compelling or unacceptable: for example, if they believe p and *if p , then q* , they will find q compelling; if they consider p and *not- p* , they will find it unacceptable to judge both. Of course, sometimes a subject might not find an instance of a rule primitively compelling or unacceptable because they’re tired, or not paying attention, or because the thoughts involved are too complex. These cases don’t raise questions about your ability to think. The troubling cases are those where there is no interfering factor, but you still don’t find an instance of the rule compelling or unacceptable.

Here, too, a capacity doesn’t have to be perfect. It can sometimes fail, even where the conditions are right. You might have the capacity to follow Non-Contradiction,

but in some cases not find a contradiction unacceptable; in these cases, you might just accept the contradictory beliefs. Your capacity might even fail a fair proportion of the time. My claim is only that it can't fail to an arbitrary degree: at some point, you don't count as having the capacity at all. When you get to this point, it's merely lucky if you find a contradiction off-putting; you're no longer sensitive to Non-Contradiction. At this point you lose your grip on what it is for two thoughts to be in tension, and so you cease to think.

This seems plausible, but it invites a worry. Broadly speaking, this sort of worry arises for any account which holds that thinking is a distinctive kind of representational activity which mature human beings engage in, discontinuous with the representational activities of other subjects who lack the requisite conceptual resources. On the one hand, thinking is supposed to be sharply discontinuous from nonthinking; on the other hand, the extent to which we tend to find instances of logical rules compelling is clearly a matter of degree. Such accounts thus face a problem about how to characterize the line between thinking and nonthinking.

Davidson (1999, 11) writes:

Both in the evolution of thought in the history of mankind, and in the evolution of thought in an individual, there is a stage at which there is no thought followed after a lapse of time by a subsequent stage at which there is thought. To describe the emergence of thought would be to describe the process which leads from the first to the second of these stages. What we lack is a satisfactory vocabulary for describing the intermediate steps.

These intermediate steps occur in a few different cases. In individual maturation: an infant lacks the conceptual resources needed for thinking, while an adult has them. When does the child begin to think? At the level of the species: human beings evolved from earlier hominids, which evolved from earlier primates. Which creature was the first to think? These cases are about the emergence of thought, but the same issue occurs with the disappearance of thought as a result of illogicality. A thinker gradually loses their sensitivity to logical rules to a greater and greater degree, until they never find the consequences of their beliefs compelling, or find a contradiction unacceptable. At what point do they cease to think?

Before addressing the problem, let me briefly mention an episode in the history of philosophy which raises similar issues. Locke and Leibniz disagreed about the consequences of the principle that nature makes no leaps, sometimes known as the

Law of Continuity. In his *Essay Concerning Human Understanding*, Locke had used the principle to argue against drawing a sharp line between human beings and other animals:

All quite down from us the descent is by easy steps, and a continued series of things, that in each remove differ very little one from the other. ... There are some brutes that seem to have as much knowledge and reason as some that are called men: and the animal and vegetable kingdoms are so nearly joined, that, if you will take the lowest of one and the highest of the other, there will scarce be perceived any great difference between them... (1689, 3.6.12)

It is a hard matter to say where sensible and rational begin, and where insensible and irrational end... The difference is exceeding great between some men, and some animals; but if we will compare the understanding and abilities of some men and some brutes, we shall find so little difference, that it will be hard to say, that that of the man is either clearer or larger. (1689, 4.16.12)

In the *New Essays on Human Understanding*, Leibniz responded that, while nature indeed makes no jumps, ‘the beauty of nature ... asks for the appearance of jumps’ (1765, 472). While intermediate species are possible, they do not exist on our planet:

although in some other world there may be species intermediate between man and beast (depending upon what senses these words are taken in), and although in all likelihood there are rational animals, somewhere, which surpass us, nature has seen fit to keep these at a distance from us so that there will be no challenge to our superiority on our own globe. ... So I believe that the stupidest man (if he is not in a condition which is contrary to nature, through illness or some other permanent defect which plays the part of an illness) is incomparably more rational and teachable than the most intellectual of all the beasts; although the opposite is sometimes said as a joke.

Leibniz argues that even the most irrational human beings pose no challenge to this thesis, as they possess capacities which no other animal possesses, even if some factor is interfering with their exercise.

Both Locke and Leibniz supposed that there is some important difference between human beings and other animals, and both struggled with how to maintain this thesis in the face of intermediate cases. Leibniz argues that the intermediate cases don’t *actually* exist, or at least not on Earth, but this seems beside the point: the worry

is just as pressing if the intermediate cases are merely possible, or extraterrestrial. Locke accepts the actual existence of intermediate cases, but doesn't offer much help with classifying them; rather, he seems to accept that these cases cast doubt on the sharp difference between rationality and arationality. Both authors presuppose that nature makes no jumps, such that intermediate cases are at least possible. But I don't want to dispute this principle here.²

4.2.1 The vagueness of *thinking*

I'm going to respond to the worry as follows. I'll begin by sharpening the worry into an explicit contradiction. Next, I'll suggest, in response to the worry, that we should accept that thinking is a vague concept. This opens up a whole range of conceptual resources for dealing with the problem. I'll briefly comment on whether accepting this vagueness might have unwelcome metaphysical consequences.

To sharpen the worry, suppose that we can order possible subjects by the degree to which they tend to find instances of logical rules primitively compelling or unacceptable, where 0 is the lowest degree of sensitivity and 100 is the highest (that is, perfect sensitivity to the rule). Then it's uncontroversial that, other things being equal,

- (1) A subject with degree 100 is thinking.

Since logic is constitutive of thinking, we will also want to claim that

- (2) A subject with degree 0 is not thinking.

For such a subject clearly lacks the requisite tendency to find instances of the rules compelling. Finally, however, it seems plausible that

- (3) If a subject with degree n is thinking, then a subject with degree $n-1$ is thinking.

But, of course, 1, 2 and 3 are inconsistent. They constitute a sorites paradox.

The problem suggests that the limit on illogical thinking which we're considering is really incoherent, because it involves commitment to 1, 2 and 3. Notice that this problem afflicts my view specifically. Someone like MacFarlane (2002), or Leech (2015), who denies that thinkers must be sensitive to logical rules to any degree, can

²Hegel denies the principle, arguing that nature does make jumps, e.g. in phase transitions: cooling water 'does not become hard a bit at a time, as if it became first like a porridge and would then gradually harden to the consistency of ice, but is hard all at once' (*Science of Logic* 322).

deny premise (2). Someone like Kimhi (2018), who requires thinkers to perfectly follow logical rules, can deny premise (3), as it is false where n is 100.

I'm going to suggest, however, that the problem merely reflects the vagueness of the concept *thinking* – something we have good reason to accept – and that nearly any viable account of vagueness can be brought to bear here as an account of the limit of illogical thinking.

To start, then, there are good reasons to accept that *thinking* is vague: the concept has several different properties usually taken to be indicia of vagueness.³ First, *thinking* displays borderline cases. Consider a subject whose representational activity is highly incoherent, but who on occasion displays some attempt to follow out the consequences of their views or to avoid explicit contradictions. Or consider the familiar difficulty in classifying young children and certain animals as either thinking or not thinking, when it's unclear whether their representational activity is sophisticated enough to count. Second, *thinking* lacks clear boundaries, for just the same reason. Recognizing this doesn't require a very theoretically committed conception of thinking, such as the one I've developed in this dissertation: just reflect on the progression from mulling over your day (clearly thinking), to drowsy, imagistic free-associating, to dreaming (clearly not thinking). Third, *thinking* is, as seen above, soritical. For all these reasons, it seems plausible that thinking is a vague concept. It's not surprising, then, that it should be hard to say just how much conformity to logic is enough to count as thinking, or that an attempt to say so precisely leads to paradox.

This is a diagnosis of the problem, but it doesn't yet offer a treatment. In my view, nearly any viable account of vagueness can be brought to bear here as an account of the limit of illogical thinking; different accounts of vagueness yield different accounts of the shape of the limit. Let me offer two examples.

First, *epistemicism* (Williamson 1994; Sorensen 2001) is the view that vague concepts have sharp boundaries, though it is impossible for us to know where those boundaries lie. Here, the epistemicist will say that there is a sharp line between thinking and nonthinking – a precise degree of sensitivity to logical rules which is required for thinking – but that we cannot know where that line is. So the right response to a borderline case is to say that the subject is either thinking or not,

³Quine (1992, 59) makes a similar suggestion about understanding. Heal (2010, 77) argues that what ties together our uses of 'thinking' is not an essence or intension but rather 'some relation back to the agreed or paradigm cases', where the relation 'may well differ from context to context'.

but that we cannot know which. This may seem puzzling, but the epistemicist can at least hold that the puzzlement is not particular to thinking; the very same puzzlement occurs in more mundane cases like tallness or redness.

Second, the *supervaluationist* will hold that there are many different admissible precisifications of *thinking*, drawing the line in different places. However, the supervaluationist introduces the notion of an ‘admissible precisification’, which reflects a way of drawing a line that respects certain constraints built in to the concept (Fine 1975). For example, if a subject with degree m is thinking, then a subject with degree $n > m$ is thinking; if a subject with degree m is not thinking, then a subject with degree $l < m$ is not thinking. If truth is truth in all admissible precisifications, then if m is a borderline case, it is neither true nor false that a subject with degree m is thinking. In other words, the right response to some of the intermediate cases may be that we should neither affirm that the subject is thinking nor affirm that the subject is not thinking.

Other accounts of vagueness could also be brought to bear here.⁴ On views other than the epistemicist one, there will generally not be a sharp line between thinking and nonthinking – which seems to accord with our intuitive judgments. I’m not going to discuss which account of vagueness is right, because the relevant considerations are those which apply to vagueness generally, rather than anything specifically to do with thinking. For present purposes, the existence of these different accounts should be enough to defuse the worry that the claim that logic is constitutive of thinking entails a contradiction in virtue of the vagueness of *thinking*.

4.2.2 The metaphysical reality of thinking

There is, however, another problem in the vicinity. What does the vagueness of *thinking* entail about the metaphysical reality of thinking? If *thinking* is a vague concept, is there still a fact of the matter about whether some subject is thinking? This problem arises if we accept an account on which vagueness is a feature of representations, rather than of the reality they represent (Russell 1923).⁵ If there are no vague properties, then there is no single property of thinking that our concept of thinking picks out. (There are properties for each of the precisifications of thinking, but none of these is the property of thinking.) As Lewis writes, ‘[T]he reason it’s vague where the outback begins is not that there’s this thing, the outback, with

⁴On vagueness generally see Keefe & Smith (1996).

⁵So it doesn’t apply to the epistemicist.

imprecise borders; rather there are many things, with different borders, and nobody has been fool enough to try to enforce a choice of one of them' (1986, 212).

If reality doesn't contain a property which is the property of thinking, then, to be blunt, thinking doesn't exist. So the claim that logic is constitutive of thinking says that logic is constitutive of something that doesn't exist. This is bad enough; it also threatens the claim I'll defend in Chapter 6, that thinking, and therefore the logical limits of thinking, have normative weight. If thinking doesn't exist, then how can the difference between thinking and failing to think matter?

In response, I want to make two points. First, if the Lewisian claim – that vagueness is a feature of representations, not reality – is read in a strong, eliminativist way, then there's good reason to resist it. Second, if it's read in a weaker, more deflationary way, then the difference between thinking and failing to think might still matter.

On the strong reading, we should eliminate vague language from serious theorizing, because the entities that vague terms purport to represent do not exist. On this reading, there really is a problem for constitutivism, because it attempts to seriously theorize about something that doesn't exist. But this reading would also have us eliminate theorizing about planets, animals and artefacts, as well as colours, bodily states and moral properties. This is just not a tenable view. If serious theorizing is this closely tied to what is metaphysically real, then we have good reason, independent of logical constitutivism, to find a metaphysics and a meta-metaphysics which allow us to hold that vague things and properties can be real.⁶ (Though I don't know what such an account should look like.)

On the weaker reading, we accept that vague objects don't exist, but we are still allowed to talk as if they did (outside the Ontology Room, as it were). In other words, metaphysical eliminativism need not entail linguistic elimination; not every usable predicate needs to correspond to a real property. I want to suggest that someone who accepted this reading could still hold that the difference between thinking and failing to think matters, even if there's no property of thinking.

If there's no property of thinking, it would make sense to accept what Field (2017) calls an *egocentric* account of thinking. On such an account, terms like 'judges' and 'infers' don't pick out a set of real properties and relations that subjects stand in to thoughts. Rather, each of us uses these terms to track our acts of translating other subjects' utterances into our own. If A makes some utterance *s* and B says, 'A

⁶Of course, there are problems for vague objects – see e.g. Evans (1978).

judges p ', this means that B is willing to translate the utterance s as meaning p in B's idiolect. But there's no objectively correct translation scheme. It's possible that B would translate A's utterance in one way, while C would translate it in a different way (or find it untranslatable), and there's no further fact making one of them right. Similarly, it might be possible that B takes A to be thinking, while C takes A to be failing to think, and there's no further fact making one of them right. This would help explain why *thinking*, like other vague predicates, has multiple correct stopping-places: while there are paradigmatic cases of thinking and paradigmatic cases of failing to think, competent users can draw the boundary between them in different places, and can recognize that other stopping-places would be equally correct (Raffman 2014, 94).

On an egocentric account, although the difference between thinking and failing to think doesn't correspond to a real property, a practice of applying or refusing to apply the predicate *thinking* might still have a point. At an individual level, if C takes A to be failing to think, then C will treat A as lacking general abilities to respond to rational relations among thoughts; at best, A is hardwired to recognize certain connections, but is unable to revise those connections. So if C takes A to be failing to think, then there are a whole range of cognitive projects that C will not engage in with A. At a social level, if the term 'thinking' has some positive normative valence, then a practice of applying or withholding the term from subjects depending on how logical they are might serve to coordinate different subjects around a common set of logical rules.⁷

4.3 The second limit on illogicality

In this section I argue for the second limit on illogical thinking:

Second Limit Thinkers must not disregard primitively compelling or primitively unacceptable judgments or inferences.

The idea here is that even if you have logical capacities, it doesn't follow that all of your representational activity involves their exercise. In some cases, you aren't exercising those capacities, and so are not thinking.

This limit marks a failure of performance, not of competence. So while a subject who violates the First Limit never thinks (at least, until they regain their logical

⁷Dogramaci (2015) argues that this is the point of applying or withholding the term 'rational'.

capacities), a subject who violates the Second Limit fails to think *on that occasion*, but they might well think on other occasions.

Whatever logic a thinker follows, the thinker's grasp of that logic will consist in finding certain judgments and inferences primitively compelling or unacceptable. The thinker cannot knowingly disregard such a compulsion. The clearest case of this phenomenon is attempting to judge an explicit contradiction, and I will take this as my example in this section. But if you think that logic sometimes permits judging a contradiction, you can substitute whichever other structures you do take logic to rule out.

The relation between this limit and the previous one is nicely brought out by Korsgaard's observation that there are two ways of failing to build a house. The first way is to try to build a good house but fall so far short that what you've built isn't a house at all. The second way is to try to build something other than a house – perhaps a non-house that looks like a house.

[The builder] may be doing one of two things. He may be guided by the norms, but carelessly, inattentively, choosing second-rate materials in a random way, sealing the corners imperfectly, adding insufficient insulation, and so on. But he may also, if he is dishonest, be doing this sort of thing quite consciously, say in order to save money. In that case, surely we can't say that he is trying to build a good house? No, but now I think we should follow Socrates' lead, and say that he is not trying to build a house at all, but rather a sort of plausible imitation of a house, one he can pass off as the real thing. (2009, 31)

Someone who isn't trying to meet the standard of a good house at all, but is trying to fall short of that standard (for whatever reason), is, Korsgaard argues, not trying to build a house at all. We can draw the same distinction among ways of failing to think. First, someone might try to follow logical rules, but be so insensitive to the rules that they violate the First Limit. Second, someone might not try to follow logical rules at all. At the times when they aren't trying to follow the rules, they aren't thinking.

This distinction maps on to another that is made in these discussions: the difference between *logical mistakes* and *logical aliens* (Nunez 2018). Someone who tries to follow logical rules but occasionally falls short is making a logical mistake. Such a person may still be thinking, if they stay within the First Limit. Someone who isn't trying to follow logical rules at all is attempting to be a logical alien – a thinker who follows rules other than the rules constitutive of thinking. They might be

trying to follow some set of rules that isn't among the rules suitable for thinking, or they might be trying not to follow any rules at all. Insofar as there can be no logical aliens, this person is not thinking. (Of course, since different sets of rules might be suitable for thinking, there's a weaker sense in which it is possible to be a logical alien: from the perspective of someone who follows intuitionistic logic, someone who follows classical logic, or some relevant logic, is a logical alien.)

4.3.1 What makes a contradiction obvious?

I've been talking about 'not trying' to follow logical rules, but what exactly does this mean? Clearly we don't ordinarily form an explicit intention to follow logical rules – for example, by representing them and making an effort to conform to them. As I argued in Chapter 3, a thinker might have no explicit beliefs at all about logical rules. So what would it be to not try to follow them?

In fact, I think the answer falls straightforwardly out of my remarks in 3.5 about the first-personal dimension of a logical capacity. I argued that a logical capacity generates primitive compulsions when a thinker is in a state covered by the rule, except where there is some interfering factor such as tiredness, inattention or the complexity of the thoughts at issue. This compulsion is a reason for the thinker to judge (or not judge) the content in question, but the reason can also be overridden by other reasons. A subject is *disregarding* a rule iff the subject does not act as the rule requires, but there is no interfering factor and no overriding reasons.⁸ This is what happens when we try to judge an obvious contradiction. An obvious contradiction, like *K committed the murder and did not commit the murder*, is, as it were, intellectually repugnant: we can't understand the content without seeing that it violates a logical rule. A subject who has the logical capacity for negation will find it primitively unacceptable to judge this content. To do so, then, involves disregarding the rule.

Attempting to judge an *obvious contradiction*, therefore, involves a failure to think. But notice that this doesn't apply to cases where interfering factors, like tiredness or inattention, spoil the operation of the capacity. In such cases, that the content is contradictory is not obvious. Whether *p*'s contradictory nature is obvious

⁸The notion of *disregarding* a logical rule needs to be spelled out such that the following sort of reasoning doesn't count as 'disregarding' Modus Ponens: *p; if p, then q; but not-q; so not-p*. This kind of reasoning clearly involves the thinker's grasp of Modus Ponens, but the thinker doesn't seem to draw a conclusion using that rule. The problem originates in Harman (1986); for discussion, see Besson (2012), and for a proposed solution, see Murzi & Steinberger (2013).

to a subject depends not only on features of p , but also on features of the subject. With this caveat, I suggest that anyone who attempts to judge such a content is, while making the attempt, not thinking.

One upshot of this limit is that while it is possible to accidentally judge a contradiction, it isn't possible to judge one whose contradictory character you are aware of. To my mind, this fits with the phenomena (cf. Sibajiban 1964; Marcus 2020, 6). On the one hand, judging some content and then realizing that it harbours a contradiction is a relatively familiar phenomenon. On the other hand, it seems impossible – even if you have good reason to believe p and good reason to believe *not-p* – to just plump for a contradiction that you know is a contradiction.⁹

To spell this out, we can borrow a distinction, drawn by Edmund Husserl in *Formal and Transcendental Logic* (1929), between *explicit* and *non-explicit* judging. We judge something non-explicitly ‘as something completely vague that comes to mind’ or ‘as the completely vague signification of a statement read, understood and believingly accepted’ (1929, 56). Husserl holds that a contradiction can be judged non-explicitly: we can, and do, assent to claims which in fact are contradictory, without being aware of this fact. But a contradiction cannot be judged explicitly. Moreover, if we make a judgment and then, upon reflection, realize that it is contradictory, ‘the underlying total belief, the one that is being explicated, forthwith and necessarily assumes the characteristic of nullity’ (58): we cannot hang on to our belief once aware that it is a contradiction. Husserl makes a parallel claim about consequence-relations: when we judge p , and come to see that q is a consequence of p , we ‘not only [judge] the consequence in fact but “cannot do otherwise”’ (1929, 189).

⁹You might worry that I’ve mislocated the source of the oddness here. Sure, if I consider a contradiction, it seems odd to decide to judge it. But that’s because I can’t *decide to judge* anything at all. I can’t decide to judge a non-contradictory thought either. In other words, you might worry that the oddness stems from assuming a kind of doxastic voluntarism. But I don’t think this is the source of the oddness. You can judge a non-contradictory content if you have evidence for it, but you can’t judge a contradiction even if you do have evidence. Suppose Inspector Morse is investigating a murder in Pitt Meadows and that K is a suspect, having committed a similar crime years ago. Some locals say they saw him walking away from Pitt Meadows on the night of the murder: this piece of evidence supports the claim that K committed the murder. But K has a train ticket showing that he left town shortly before that time: this piece of evidence supports the claim that K did not commit the murder. Now there are various things Morse might conclude, but one thing he can’t conclude that K did and did not commit the murder. But the problem with this is not that he lacks evidence for both conjuncts. It is that the conjunction is a contradiction.

4.3.2 *Reductio* proofs

You might worry that contradictions play an important role in reasoning, for example in proofs by *reductio* (Quine 1948). But surely a non-thought can't play any role in a proof. To be clear, however, I'm not claiming that there are no contradictory contents – only that such contents cannot be judged when their contradictory nature is obvious. Moreover, *reductio* proofs actually rely on the primitive unacceptability of contradictions. If contradictions weren't unacceptable, then the fact that p entails both q and *not-q* would be no reason for disbelieving p . We could just accept the contradictory consequences.

In *reductio* proofs, the aim is not to judge the contradiction but to see that it follows from certain suppositions, such that one of the suppositions has to be rejected. Coming to see this does involve exercising your logical capacities, but it doesn't involve disregarding the unacceptability of a contradiction. At most, it requires entertaining a contradiction within the scope of a supposition, without any belief-like commitment.

4.4 Can we be certain someone isn't thinking?

I've argued that there are two limits on illogical thinking. First, thinkers have to tend to find instances of logical rules compelling, and second, they can't disregard such compulsions. Now, you might worry that recognizing such limits would have a deleterious effect on rational discourse. They seem to give us license to dismiss subjects who seem to make logical mistakes, or whose inferences we can't follow, as failing to think and therefore not worth the time.

I think this is a fair concern. In response, I want to emphasize the gap between the metaphysical question of whether someone is failing to think and the epistemic question of whether we can be certain that they're failing to think. These questions come apart in principle, not just in fact. As Moore (2003a, 186) writes, 'the claim that something is nonsense is always empirical [and] provisional'.

First, the facts that determine whether someone is failing to think are not directly manifest in their utterances, behaviour, or any other evidence available from the outside. Whether someone violates the First Limit depends on whether they tend to find certain inferences compelling, and whether they violate the Second Limit depends on whether they disregard such feelings; neither of these facts can be ascertained with

certainty from outside. For example, suppose A asserts a contradiction. We can't tell with certainty whether (i) A didn't find it unacceptable because they lack logical capacities; (ii) A didn't find it unacceptable because they were tired or inattentive or didn't understand the thoughts involved; (iii) A did find it unacceptable, but decided upon theoretical reflection that this contradiction had to be true; or (iv) A did find it unacceptable, and disregarded this feeling. Only in cases (i) and (iv) is A failing to think.

Second, we can't tell with certainty which rule someone is following, for reasons that are essentially those of Kripke (1982). For any finite set of inferences a third party observes, there will be many different rules that validate those inferences and diverge elsewhere. So if all we have to go on is a finite set of inferences, we can't tell with certainty which rule someone is following. As a result, we can't tell with certainty whether they're following one of the sets of rules that is suitable for thinking. (To be clear, I don't agree with Kripke that there's no fact of the matter about which rule they're following – it's just that we can't be sure.)

Third, it's not as if we begin with knowledge of which thoughts someone is thinking and then ask which rules they're following; rather, we interpret their pattern of activity as a whole. So if someone regularly utters sentences that ordinarily mean *p*, *if p, then q* and *not-q*, we might interpret them as meaning something other than *p* or *q* in one of these sentences, such that the triad is no longer inconsistent. Interpretive charity will often favour a local reinterpretation over the global result that someone is failing to think.

Part II

Consequences of constitutivism

5 Nonsense: a user's guide

Note. This chapter is a bit of an interlude from my main line of argument. It develops a general account of what happens when we fail to think because of some semantic defect. Indirectly, the chapter responds to a worry that the earlier discussion likely raised. In Chapters 1 to 4, I argued that logic is constitutive of thinking, and defended two constraints on how illogical a subject's representational activity can be while still counting as thinking. One upshot is that sometimes we fail to think because our activity is insufficiently logical. In this chapter, I explain how the failure to think can seem like thinking – in particular, how it can include phenomena such as reasoning. This chapter is, however, written as a standalone piece; it doesn't presuppose the views defended elsewhere in the dissertation, and applies to a broader class of failures to think than those due to illogicality alone.

5.1 Introduction

Many philosophers suppose that we can think we are saying or thinking something meaningful when in fact we're not saying or thinking anything at all: we are producing nonsense. But what is nonsense?

Early twentieth-century philosophy of language abounded in allegations of nonsense. Russell (1919) suggested that violations of logical type-restrictions result in nonsense. Wittgenstein (*TLP* 6.54) held that much philosophical discourse, including most of his *Tractatus Logico-Philosophicus*, was meaningless. Ayer (1936) argued that verifiability is a semantic constraint, such that metaphysical language is meaningless. Ryle (1949) held that when we make a category mistake, we produce nonsense.

But the idea is still around. Here are two more recent uses of it:

[J]ust as we cannot know a priori or with Cartesian certainty whether any particular thing we think or say is true, so we cannot know a priori or with Cartesian certainty that in seeming to think or talk about something we *are* thinking or talking about – *anything at all*. We cannot know a priori *that we mean*. (Millikan 1984, 10)

Look, it can be true that people think they have a thought when they don't. They can be deeply attached to a linguistic formulation that upon reflection doesn't say anything. ... So a lot of our thoughts are really not thoughts, they're things masquerading as thoughts. (Appiah 2017)

Rather than giving a definition of nonsense, I'll take the concept to be fixed by its theoretical role. 'Nonsense' is a term of criticism. It picks out a failure to express, or have, a thought: a failure more basic than saying or thinking something significant but false (Routley 1969; Goddard 1970). It wouldn't be a useful term of criticism, however, if failures to make sense were always obvious. Rather, it's implicit in the way philosophers have used the concept that something can appear to make sense even though it does not. This is why many twentieth-century philosophers took it as their job to expose latent nonsense.

The concept of nonsense is closely related to the theory of meaning. The theory of meaning specifies the facts in virtue of which a sentence, as used by a particular speaker or community, has a given content. It thereby entails conditions for saying something meaningful at all; when we violate one of these conditions, we produce nonsense (Cappellen 2013, 25). We might wonder, however, whether meaning is the sort of thing that is suited to being theorized at all. If we experience something as meaningful, how can theory tell us it is meaningless? I'm not going to provide a defence of theorizing about meaning here.¹ But I would note that the concept of nonsense doesn't require us to exclude experience altogether: we can allow that meaning is ordinarily something that shows up in experience (Zwicky 2019). What we cannot say, if we are to use the concept of nonsense critically, is that the experience of meaningfulness is sufficient for meaning. Theorizing can sometimes show that an apparent thought was really nonsense.

In this paper I'm not going to argue that we should use the concept of nonsense in the way twentieth-century philosophers did; nor will I take a position on which cases count as nonsense.² Rather, I'm interested in what nonsense would have to be, for the project of those philosophers to be a viable one. I'm interested, in other words, in what would be required of an account of nonsense for the concept to do critical work. In my view, such an account needs to explain two things:

1. How nonsense can appear meaningful, and
2. How nonsense is in fact not meaningful.

These requirements lead to two constraints on an adequate account of nonsense. The

¹For discussion of the idea that meaning is what a theory of meaning is a theory of, see Dummett (1975) and McDowell (1976). For criticism, see Wright (1981).

²So I will not be arguing against philosophers who suggest that many cases of apparent nonsense – such as category mistakes – are really just necessary falsehoods (Pap 1960; Goldstick 1974; Camp 2004; Magidor 2013); for responses, see Routley (1969) and Glock (2015, s. 7).

Engagement Constraint, discussed in 5.2 below, is that nonsense can be used in certain ways. The **Austerity Constraint**, discussed in 5.3, is that nonsense does not express thoughts. Some extant accounts fail, I will argue, because they do not meet one constraint or the other. In 5.4, I offer an account which meets both constraints and respond to some objections. Finally, in 5.5 I conclude with a hard question raised by nonsense.

5.2 The engagement constraint

In this section I'll defend the following constraint on accounts of nonsense:

The Engagement Constraint Nonsense can be engaged with.

'Engagement' is a loose term for various ways in which we seem to use nonsense, not just mention it, bringing to bear our ordinary conceptual capacities. As Annette Baier writes (1967, 520), nonsense 'need be neither useless nor lacking in order and discipline'. We need to account for engagement if we are to explain the appearance of sense that nonsense can present. In particular, there are two main ways of using nonsense that we have to explain: reasoning with nonsense and embedding nonsense within that-clauses.

5.2.1 Reasoning with nonsense

We draw apparent inferences from nonsense. Different instances of nonsense will appear to license different inferences, depending on the concepts involved. I'll illustrate this point with two examples: a version of the Liar Paradox and a category mistake.³

So suppose the following sentence is nonsense:

- (1) 1 is false.⁴

To someone under the illusion that 1 expresses a thought, it will seem that 1 entails

- (2) '1 is false' is false.

³Some deny that semantic paradoxes and category mistakes result in nonsense (see previous note). For my purposes, nothing hangs on this, as long as it is conceded that there exists nonsense that looks sufficiently sentence-like to underwrite 'inferences' like the one in the text. Any philosopher who wants to use 'nonsense' as a term of criticism must think that such instances exist.

⁴For accounts on which the Liar expresses no proposition, see Kripke (1975) and Rumfitt (2014).

Now consider a different instance of nonsense:

(3) Goodness is hexagonal.

Someone who thinks that 3 makes sense will think 3 entails

(4) Something is hexagonal.

By contrast, someone who thinks that 1 makes sense would be 'wrong' to think that 1 entails 4. We have here a notion of 'correctness in inference' for nonsense.

Let us say that someone who reasons from 1 to 2, or from 3 to 4, is *reasoning with* the relevant instances of nonsense. Then our reasoning with 1 differs from our reasoning with 3. Of course, supposing 1 and 3 really are nonsense, these 'inferences' are defective. Nonsense doesn't entail anything.⁵ But an account which fails to capture these sorts of moves is missing a crucial way we engage with nonsense, drawing on the conceptual and linguistic capacities we exercise in ordinary speaking and thinking.

Moreover, reasoning with nonsense can be central to our recognition of it as nonsense. This is characteristic of the critical use of the concept of nonsense by philosophers like Russell and Ryle (Goddard 1970, 12). Consider Russell's discussion of existence. Russell held that existence was a second-order predicate applicable to first-order predicates. He denied that there was any first-order equivalent (1919, 206):

As regards the actual things there are in the world, there is nothing at all you can say about them that in any way corresponds to this notion of existence. It is a sheer mistake to say that there is anything analogous to existence that you can say about them. You get into confusion through language [...]. I mean, it is perfectly clear that, if there were such a thing as this existence of individuals that we talk of, it would be absolutely impossible for it not to apply, and that is the characteristic of a mistake.

To show that a certain concept doesn't exist, Russell supposes that it does exist and considers how it would apply. That is, Russell notes that, if there were a first-order existence predicate, then for any object *o*, it would be true to say that *o* exists using this predicate. And Russell supposes that a genuine predicate cannot be such that it must be true of everything: if you can't be wrong, then you can't be right either; so there is no first-order existence predicate. But of course, if Russell is right about

⁵Assuming that the notion of entailment is semantic and not merely syntactic.

this, then there is no such thing as the thought that o exists, and therefore no such truth about o .

Here is another example. Recall

- (1) 1 is false.

In reasoning with 1 I may start by supposing that 1 is false. Then I infer that, since 1 says that it is false, and it *is* false, 1 is true. Of course, if 1 is true, then, since it says that it is false, 1 is false. Now I know I'm in trouble, as 1 seems to be true if and only if it is false, but no sentence can be both true and false. So I conclude that 1 does not express a thought at all: it is nonsense, and is therefore neither true nor false and does not entail anything. Again, reasoning with 1 was crucial to recognizing the problem.

5.2.2 Nonsense-attributions

This brings us to the second way of engaging with nonsense. Once we recognize that a given thought is illusory, or that a given sentence doesn't express a thought, we can express our recognition of this fact using a that-clause (Sorensen 2002; Cappellen 2013).

The kind of thing I mean is best brought out with an example. Suppose that I've been reading Wittgenstein's *Tractatus* and am gripped by the Picture Theory of Meaning. Yes, I say: a picture really is a fact! (*TLP* 2.141) But as I progress through the book, I begin to suspect that its propositions are meaningless by its own lights. Wittgenstein, I think, was right: 'anyone who understands me eventually recognizes them as nonsensical' (6.54). I express my realization as follows:

- (5) It is nonsense to say that a picture is a fact.

Call such sentences *nonsense-attributions*. They seem to use rather than mention the phrase in the that-clause.⁶ For example, it seems that we could not understand 'It is nonsense to say that goodness is hexagonal' without having the concept of a hexagon. In general, the cognitive value of a nonsense-attribution, the rational difference that accepting it makes (Frege 1892), depends on understanding the concepts used within

⁶I'm aware that nonsense-attributions sound odd. Perhaps this is better: 'To say "A picture is a fact" is nonsense'. I'm happy to grant this, as long as it is clear that this locution can be read in a non-metalinguistic way, where understanding the whole requires understanding the double-quoted words.

the that-clause.

Some might argue that nonsense-attributions are metalinguistic – that the nonsense is merely being mentioned. I don't deny that they can be read in a metalinguistic way, but there is also a non-metalinguistic reading.

On the metalinguistic reading, a sentence of the form 'It is nonsense to say that *S*' says the same as one of the form 'The string *S* does not express a thought'. On this reading, the cognitive value of a nonsense-attribution is that a particular sequence of symbols does not express a thought. Note that it's possible to appreciate this without being able to understand any element of the sequence of symbols: for example, I might learn that a particular sequence of letters in shorthand script doesn't say anything, even though I can't read shorthand.

On the non-metalinguistic reading, we need to understand the concepts used within the that-clause to understand the nonsense-attribution. To see that such a reading is available, let's return to the *Tractatus* example:

- (5) It is nonsense to say that a picture is a fact.

Now contrast this with another sentence:

- (6) It is nonsense to say that das Bild eine Tatsache ist.⁷

If you don't speak German, you won't be able to engage with the content of the that-clause in 6, while you will be able to engage with its content in 5. For example, you won't be able to identify the concepts used within the that-clause. For this reason, 5 is informative in a way that 6 is not. However, on the metalinguistic reading, both 5 and 6 point out that certain strings do not express thoughts. The reading on which 5 says more than 6 must consider nonsense-attributions as more than metalinguistic.⁸

You might wonder how nonsense-attributions work semantically. For present purposes, I'm saying nothing about this: the point is simply that any viable account

⁷You might object that 8 is ungrammatical (Quine 1961, 135). That might be right. But it would provide further evidence that nonsense-attributions are not covertly metalinguistic, as 'It is nonsense to say "das Bild ist eine Tatsache"' is unquestionably grammatical.

⁸Someone who favours the metalinguistic view might point out that they know more about 'A picture is a fact' than about 'Das Bild ist eine Tatsache': they know the meanings of the words in the former string. But on the metalinguistic view, this knowledge plays no part in understanding 5. So 5 is not more informative than 8. At best, 5 can be combined with other knowledge to draw inferences that cannot be drawn from 8. My point is that 5 is in itself more informative than 8 for our Anglophone *Tractatus* reader.

of nonsense ought to validate the intuition that nonsense-attributions are informative in a non-metalinguistic way. This intuition doesn't dictate the semantics we give. Later in the paper, when I give my positive account of nonsense, I will explain what is going on in nonsense-attributions.⁹

I've described two ways we seem to use nonsense rather than only mentioning it: first, we reason with nonsense, and second, we make nonsense-attributions. To account for these phenomena, we must satisfy Engagement. I'll now show that a prominent account fails to satisfy the constraint.

5.2.3 The minimalist account

The *minimalist account* says that, when we produce nonsense, we produce mere words which fail to bear meaning: at the level of language, there is a string of symbols, but at the level of thought, there is nothing. A.W. Moore takes this position (2000, 198-9):

...there is one and only one way in which an utterance can fail to be a representation, namely because the words involved in it have not been assigned suitable meaning: they have not been assigned such meaning as would give the utterance content.

Later, he writes that even the most apparently significant nonsense, like the philosophical nonsense we produce 'when trying to express the inexpressible', is 'none other than the the pure and utter nonsense of 'Phlump jing ux.'

This position is often thought to follow from Frege's Context Principle, which says that words have meaning only in the context of a sentence, or that concepts can occur only in the context of a thought (Frege 1884, x). Supposing that some sentence does not express a thought, the suggestion is that the elements of the sentence cannot express any concepts either.

I'm going to argue that the minimalist account fails to satisfy Engagement. After this, I'll show that the Context Principle does not, in fact, support this account. The basic problem is that, on the minimalist account, when we produce nonsense we aren't using our conceptual capacities. But our engagement with nonsense draws on precisely these capacities.

⁹Nonsense-attributions are related to another type of sentence in which nonsense is used within a that-clause, e.g. 'Wittgenstein believes that a picture is a fact'. It's disputed whether such a sentence makes sense (Diamond 2000, s. 4; Sorensen 2002), but there's some intuitive pressure to say it does. While I won't have room to discuss this, I believe my account can be extended to deal with such sentences.

Let's take the two forms of engagement in turn. First, recall the move from 3 to 4:

- (3) Goodness is hexagonal.
- (4) Something is hexagonal.

If the words in 3 don't bear any meaning, but the words in 4 do bear meaning, then the string 'hexagonal' in 3 doesn't have the same meaning as it does in 4. But any account of the move from 3 to 4 must invoke this common feature of the two sentences. If 'hexagonal' doesn't have the same meaning in 3 and 4, then it's unclear why moving from 3 to 4 is better than moving from 3 to, say, 'Something is square'.¹⁰ More generally, reasoning with nonsense relies on the presence of meaningful words and structures in nonsense. So the minimalist account cannot explain reasoning with nonsense.

Second, recall the difference between the nonsense-attributions 5 and 6:

- (5) It is nonsense to say that a picture is a fact.
- (6) It is nonsense to say that das Bild eine Tatsache ist.

If the words within the that-clauses in 5 and 6 do not bear any meaning, then the only difference between 5 and 6 is that they contain different strings of letters. But then we're forced to adopt the metalinguistic reading of nonsense-attributions: what 5 and 6 tell us is that certain strings lack meaning. And we're unable to explain why 5 seems to tell an English-speaking reader more than 6 does.

For these reasons, the minimalist account does not satisfy Engagement.

Now, all of this heavy weather about the minimalist account's defects might seem unnecessary. After all, the proponents of the account don't argue that we should accept it because it preserves the appearances. Rather, they argue that we should accept it because it follows from an important principle about meaning. Here is Moore (2000, 199):

The guiding principle here is that there cannot be as it were positive semantic reasons for an utterance's failing to make sense. It cannot be because of what the parts of the utterance do mean, that the whole thing does not mean anything. The meaning of the parts is their contribution to

¹⁰The fact that the same string, 'hexagonal', occurs in both 3 and 4 is insufficient, given the possibility of homonyms.

the meaning of a range of wholes. ... The only thing about a word that can prevent utterances in which it is used from being proper representations is its not having any meaning at all.

Cora Diamond makes the same case in a different way. Supposing that the meaning of a term is its contribution to the meanings of sentences, then if a sentence is meaningless, its constituent terms must also be meaningless. So 'a sentence which does not make sense does not contain words which can be said to mean what they do elsewhere' (1981, 21; cf. Diamond 1988; Conant 2002). Now, if the meaning of a term is essential to its identity, then no term in a meaningless sentence is the same as a term in a meaningful sentence (1981, 11):

If I know the rules of the language, I know what a sentence composed in such-and-such a way out of such-and-such Elements says; but I do not know (there is no such thing as knowing) that what I see or hear is this Element, unless the whole of which it is part has a sense to which the meaning of this Element contributes in the way determined by the rules.

In both formulations, the argument depends crucially on the following principle:

CP The meaning of a term is its contribution to the meanings of sentences.¹¹

I'm going to argue that there are two ways of reading CP. On the first reading, the meaning of a term is its contribution to the meaning of the particular sentence in which it occurs. This reading of CP is implausible, but does support the minimalist account. On the second reading, the meaning of a term is its contribution to the meaning of sentences in general in which it occurs. This reading of CP is plausible, but doesn't support the account.¹²

On the first reading, the meaning of a term is its contribution to the meaning of the particular sentence in which it occurs. For example, the meaning of 'hexagonal' in 4 consists in its contribution to the meaning of 4. On this reading, we cannot ask for the meaning of a term in general – we cannot, for example, ask what contribution 'hexagonal' makes to sentences of the form '*a* is hexagonal'. Rather, we can only ask what contribution a term makes in a particular sentence.

This does support the minimalist account. For suppose that 3 has no meaning. Then there is nothing to be said about the contribution that 'hexagonal' makes in its

¹¹To account for context we should say 'utterances', but for present purposes we can ignore this.

¹²Glock (2015, s. 5) responds to the same argument in a different way.

occurrence in 3. The fact that 'hexagonal' does contribute to the meaning of other sentences, like 4, is irrelevant, given that meaning is particular rather than general.

The problem is that on this reading CP makes it unclear how linguistic understanding is possible. We learn the meaning of a word from a finite set of sentences. If there's no general contribution that a word makes to each sentence of which it is a part, then it's unclear how we could ever understand what a word means in a sentence we haven't encountered before. In Diamond's terms, we could never be sure that the terms in a new sentence are the same as terms we already know.¹³ So we should reject this reading of CP.

On the second reading, the meaning of a term is its contribution to the meaning of sentences in general in which it occurs. As Diamond puts it – stating the view she rejects – 'it is the *general* possibility a word has of contributing to sense that confers meaning on it' (1981, 18). Since meanings are general, we can learn the meaning of a word from some finite number of instances and apply this understanding in new cases.

This reading doesn't trivialize CP. CP tells us that if we don't understand the contribution a term makes to sentences, then we don't understand the term. For example, suppose I have a sophisticated theory about some object which I associate with the sign \neg . Nevertheless, if I don't understand that $\neg p$ is true when p is false, then I don't understand the meaning of \neg .

Read this way, however, CP doesn't support the minimalist account. The meaning of 'hexagonal', for example, is the general contribution it makes to the meaning of sentences. When the term occurs in 3, its meaning remains the same, because it doesn't depend on the meaning of the particular sentence in which it occurs. The lack of meaning of the whole is no reason to deny that the parts have meaning.

Still less does it follow that, as Moore says, 'It cannot be because of what the parts of the utterance *do* mean, that the whole thing does not mean anything.' Suppose that 'goodness' refers to a normative property and 'hexagonal' ascribes a property applicable only to extended things. Then the reason why 'goodness is hexagonal' is meaningless is precisely because of what its parts do mean. Both 'goodness' and 'hexagonal' have meanings, but the latter is not defined for the kind

¹³Indeed, on Diamond's view it's unclear how we can ever understand a new sentence. We can't know what its terms mean until we know what the sentence means. But how do we know what the sentence means, unless we know what its terms mean? Diamond writes (1981, 21): '*That* such-and-such a word is a working part of a sentence, and that it is *its* content we must grasp to understand the sentence ... cannot be told by observation.' But she doesn't say how it can be told.

of thing referred to by the former.¹⁴

I've argued that the minimalist account fails to satisfy Engagement. I've also shown that CP does not support the minimalist account. But I don't want to leave the impression that the minimalist account is unmotivated. Rather, its deepest motivation comes from difficulties about nonsense at the level of thought.

Consider a remark Wittgenstein makes in the Preface to the *Tractatus*: 'to be able to draw a limit to thought, we should have to find both sides of the limit thinkable (i.e. we should have to be able to think what cannot be thought). It will therefore only be in language that the limit can be drawn'. Wittgenstein is pointing to an asymmetry between language and thought: while there's no difficulty in explaining why a sentence is meaningless without presupposing that it is meaningful, there is some difficulty in explaining why a 'thought' is meaningless without presupposing that it is meaningful.

For suppose we say that, when a subject attempts to understand a nonsense sentence, they exercise the same capacities they would exercise in understanding the sentence's words and structure elsewhere. It's unclear, then, why a subject attempting to understand nonsense fails to grasp a thought: the subject's activity looks the same as in the good case. It looks like any substantive explanation of why a given attempt at thinking failed will characterize it in ways that assimilate it to successful thinking. The motivation for the minimalist account is that it precludes any such substantive explanation.¹⁵

I'm not going to respond to this issue here. I'll respond to it later on by giving an account which allows for such substantive explanations without assimilating nonsense to sense at the level of thought.

5.3 The austerity constraint

In this section I defend a second constraint on accounts of nonsense:

The Austerity Constraint Nonsense does not express thoughts.

In a nutshell: if nonsense expresses thoughts, it can play the same role as sense. To

¹⁴I'm not claiming that this kind of explanation (sometimes called a coupling theory (Routley 1966, 180)) explains all cases of nonsense.

¹⁵Compare Conant (1992), who similarly rejects any substantive explanation of why logically alien thought is impossible.

the extent that nonsense can play the same role as sense, the concept of nonsense loses its critical force. So nonsense must not express thoughts.

5.3.1 Nonsense thoughts

To approach the Austerity Constraint, let's consider an account which directly violates it: the *nonsense thoughts* account, on which nonsense expresses genuine but defective thoughts.¹⁶ This account easily satisfies Engagement. Since a nonsense thought is a thought, it can form part of chains of reasoning. A nonsense thought can also fall within a that-clause, as in nonsense-attributions.

Despite these benefits, nonsense thoughts are a dead end. They make nonsense too much like sense. Consider two important theoretical roles played by thoughts. First, thoughts explain a central kind of understanding: I've understood a sentence if I grasp the thought it expresses. Second, thoughts explain a central kind of communication: I've successfully communicated with you if you grasp the thought I intended to express (Grice 1957). Now, if there are nonsense thoughts, then both roles can be played by nonsense as well as by sense. Suppose that 3 expresses the nonsense thought that goodness is hexagonal. I understand 3 if I grasp this thought; if I grasp a different one, I've failed to understand 3. If I utter 3 to you and you grasp this thought, then I've used 3 to successfully communicate. But if a nonsense thought can do the same work as an ordinary thought, then the concept of nonsense loses its critical force. That a given area of discourse is nonsense becomes an evaluatively neutral fact, not a defect.

In response, the proponent of nonsense thoughts should explain how nonsense thoughts are defective compared to ordinary thoughts, blocking the assimilation of our engagement with nonsense to our engagement with sense. I consider such a response in the next section.

5.3.2 What's wrong with nonsense?

The natural way to spell out the defectiveness of nonsense is to hold that nonsense thoughts are not truth-apt.¹⁷ The proposal is that nonsense thoughts exist but lack truth-value – they are neither true nor false – while ordinary thoughts have a

¹⁶Sorensen (2002) argues, relatedly, that there are meaningless objects of belief (albeit statements rather than propositions).

¹⁷I'm going to talk about sentences as having truth-values, abstracting from the role of context.

truth-value. As truth is an important aim of rational discourse (Dummett 1973, ch. 10; Priest 2006, s. 4.5), this is reason to avoid speaking nonsense.

To see why this won't work, recall an argument from Dummett (1959), which forces us to be more precise about the sense in which nonsense is neither true nor false. Dummett is responding to Strawson's account of empty singular terms, like 'the King of France'. Strawson had suggested that statements like 'The King of France is bald' are meaningful – we know how the world would have to be for them to be true – yet neither true nor false (1950, 330). Dummett argues that this is untenable. If the statement is meaningful, it can be used to make an assertion, and if the world is not the way it would have to be for the assertion to be true, then the assertion is false. Of course, such statements might display odd compositional behaviour – 'The King of France is not bald' is also false – but, Dummett argues, their normative status is the same as any other falsehood.

Similarly, if nonsense expresses thoughts, then it can be used to make assertions; then, if the world is not the way it would have to be for the assertions to be true, the assertions are false. An assertion of nonsense might display odd compositional behaviour, as its negation will typically also be nonsense; but its normative status is the same as any other falsehood. But nonsense was supposed to be a distinctive normative status.

We can set out the objection more precisely using Dummett's distinction between *assertoric content* and *ingredient sense* (1973, ch. 12).¹⁸ These levels of content are associated with different truth-values. Assertoric content is what is expressed by a sentence in assertion. Dummett maintains that the point of assertion is to exclude possibilities: in asserting p , we indiscriminately exclude all possibilities other than those in which p holds. Either the actual world falls into the excluded class, or it does not; so at the level of assertoric content, every assertion is either true or false.

Ingredient sense, by contrast, tracks the contribution a sentence makes to complex sentences of which it is a part. Different sentences may have the same truth-value at the level of assertoric content, but behave differently when part of complex sentences. For example, consider the sentences a is F and b is F , and suppose that a is an empty name and b names an object that doesn't fall under F . Then at the level of assertoric content, a is F and b is F are both false. But a is not F is also false, while b is not F is true. Negation is sensitive to a distinction among kinds of falsity which doesn't show up at the level of assertoric content. In Dummett's terminology,

¹⁸For further discussion, see Suszko (1977) and Shaw (2014).

a is F is false at the level of assertoric content but neither true nor false at the level of ingredient sense.

Now, the proponent of nonsense thoughts must specify at what level nonsense thoughts are neither true nor false. Consider someone who asserts 3. We're supposing that the speaker succeeds in saying something – namely, that goodness is hexagonal. The speaker succeeds in expressing a certain kind of content – the content characteristic of declarative sentences. Some possibilities are excluded by this content. In fact, all of them are. So the actual world is excluded. So the assertion of 3 is simply false, at the level of assertoric content.

However, 3 will interact differently with negation than an ordinary falsehood. Suppose someone asserts

(7) Goodness is not hexagonal.

There's a reading of 7 on which it is false just as well as 3, and for just the same reason.¹⁹ This compositional pattern is what it means for 3 to be neither true nor false at the level of ingredient sense.

But this is insufficient to capture the defectiveness of nonsense. Assertoric content is the notion bound up with the rational purpose of assertion: it captures the success or failure of an assertion in light of that purpose. Ingredient sense, by contrast, is a technical device serving to capture compositional effects. Truth-values assigned at the level of ingredient sense have no normative weight.

So to say that the assertion is neither true nor false is not to say that it fails more badly than a false assertion, but only that it is a false assertion of a particular sort, displaying particular compositional behaviour. The assertion is bad only because, at the level of assertoric content, it is false. Nonsense, however, is supposed to be a failure more basic than the failure to say something true. So a treatment of nonsense as expressing thoughts that are neither true nor false only at the level of ingredient sense obviates the critical force of the concept of nonsense.²⁰

Nonsense, therefore, must lack assertoric content. To maintain this, we must satisfy Austerity: we must deny that nonsense expresses thoughts, rather than holding that the thoughts it expresses are neither true nor false.²¹

¹⁹There is also an 'external' reading of 7 on which it is true. But in ordinary cases the two readings don't come apart (Routley 1966, 181).

²⁰For additional arguments against assimilating nonsense to falsity, see Routley (1969) and Goddard (1970).

²¹This is consistent with using a three-valued logic to represent the interaction of nonsense with

5.4 A pretence account of nonsense

I've set out two constraints that an account of nonsense must meet. It's hard to meet both constraints, because they pull in opposite directions. The nonsense thoughts account satisfies Engagement, but fails to satisfy Austerity. The minimalist account satisfies Austerity, but not Engagement. In short: it's hard to explain our dealings with nonsense without assimilating nonsense to sense.

This tension is not only theoretical. We can see how it shows up in experience in Diamond's description of trying to understand someone who utters nonsense (2000, 157-8):

When you understand someone who utters nonsense, you are not, on the one hand, remaining as it were outside his thought and describing what goes on from the point of view of empirical psychology. But, on the other hand, you are not inside his thought as you are when he makes sense and you understand what he says, because there is no such internal understanding, there is no thought that such-and-such to understand. ... There is, as I said, no inside. But what it is to understand a person who utters nonsense is to go as far as one can with the idea that there is.

For Diamond, this is how we should read the *Tractatus*: by imaginatively entering into the illusion that its sentences express thoughts.

In this section I offer an account of nonsense in the spirit of Diamond's suggestion that engagement with nonsense is an exercise of imagination. First, I explain what semantic pretence is.²² Next, I introduce some background about understanding and logical form. Finally, I draw on this background to give the account, showing that it satisfies both constraints.

5.4.1 Pretence and make-believe

Pretence accounts codify games of make-believe, of the sort that children play. A game of make-believe will typically involve some really existing items (*props*), about which something is expressly pretended to be the case (*initial stipulations*), and principles for generating further content in the pretence (*principles of generation*)

sense. Kripke, for example, uses the third value for sentences that 'do not express propositions' (1975, 701). See n. 30 below.

²²The classic source for pretence accounts is Walton (1990; 1993). For a thorough recent discussion, see Armour-Garb & Woodbridge (2015). My view differs from that of Armour-Garb & Woodbridge in using pretence to explain our engagement with nonsense, rather than to explain the functioning of ordinary, successful language.

(Armour-Garb & Woodbridge 2015, 39ff.). Consider, for example, a game about a bank robbery. The props are a child, X, and some newspapers. The initial stipulations are that X is a bank robber and the newspapers are cash. But the game's content is not fixed solely by these stipulations (Crimmins 1998, 5). Real-world facts – in particular, facts about the props and their relations – can be incorporated into the pretence. Thus, when X bolts with an armful of newspapers, in the pretence the bank robber is making a getaway.

For present purposes, we're interested in *semantic pretence* – pretence about the meaning of certain sentences or purported thoughts. The props are words, concepts and mental events; the initial stipulations assign meanings to these items which they ordinarily lack. In this way semantic pretence can increase the expressive resources of our language without increasing our stock of words.

Recent philosophical history offers some reason to think that a semantic pretence account will meet our two constraints. Late in *The Varieties of Reference*, Evans sought to give an account of sentences containing empty names, like 'Sherlock Holmes does not exist' (1982, ch. 10). On the one hand, Evans held that empty names are genuinely used, not just mentioned, in such sentences. Assuming that using a name requires knowing what it refers to, this explains why you can't understand 'Sherlock Holmes does not exist' without knowing who Sherlock Holmes is. On the other hand, Evans maintained that, since the sense of a name is a way of presenting its bearer, an empty name has no sense. It follows that sentences containing empty names do not express thoughts. Evans navigated between these two requirements with a semantic pretence account of the use of empty names. Insofar as Evans' two requirements look a lot like the Engagement Constraint and the Austerity Constraint, we have reason to think that a semantic pretence account will be adequate to nonsense more generally.²³

Before considering nonsense, it'll be useful to begin with an example where pretence operates on an otherwise meaningful sentence. Let's stay with the bank-robbery pretence. I'll say that a sentence is 'make-believedly true' when an assertion of it would be correct within the pretence (Evans 1982, 354ff.; Crimmins 1998, 15). The basic rules of the pretence, then, are as follows.²⁴

²³The connection between Evans' pretence theory and nonsense is drawn by Moore (2003a, 188).

²⁴I use *P*, *Q* and *R* as variables running over sentences and *P*['a'/'b'; 'c'/'d'; ...] for the result of replacing 'a' with 'b', 'c' with 'd' and so on in *P*.

Bank robbery pretence²⁵

Props. The props are the terms 'The bank robber', 'cash' and sentences containing those terms.

Initial Stipulations. For a sentence P which contains one or more of 'The bank robber' and 'cash', P is make-believedly true if and only if P ['The bank robber'/'X'; 'cash'/'newspapers'] is true.

We also have two more general rules (Evans 1982):

Principle of Generation. If P is true, and if there is no set of make-believedly true sentences $Q_1 \dots Q_n$ such that if $Q_1 \dots Q_n$ were true then P would not be true, then P is make-believedly true.

Recursive Principle. If $P_1 \dots P_n$ are make-believedly true and if $P_1 \dots P_n$ were true R would be true and there is no set of make-believe truths $Q_1 \dots Q_n$ such that if $Q_1 \dots Q_n$ were true then R would not be true, then R is make-believedly true.

These rules specify the set of make-believedly true claims. Some claims are expressly made-believe, while others are determined by a function from real-world truths to make-believedly true claims. This makes it possible to recover real-world content from make-believe content, or vice versa (Yablo 2014, s. 10.2).

Now consider an assertion of 'The bank robber dropped the cash while fleeing.' This assertion might lack a truth-value, as phrases like 'The bank robber' and 'the cash' may lack reference. But given the rules of the pretence, we know what has to be the case for the assertion to be make-believedly true: the assertion is correct, within the pretence, if and only if X dropped the newspapers while fleeing.

5.4.2 Logical form and partial understanding

A pretence account of nonsense raises special difficulties because the participants pretend not only that something is the case, but that something makes sense. To give such an account, we need some background about how we understand thoughts and sentences.

I'm going to suppose that recognition of logical form is essential to our understanding of thoughts and sentences. By 'logical form', I mean the way a thought or

²⁵I'm borrowing the helpful format of Armour-Garb & Woodbridge (2015).

sentence is composed of its elements. I'll speak of a sentence as 'having' or 'being of' a particular logical form, but this shouldn't be taken to imply that a given sentence can have only one form (Oliver 2010). For example, 'If Hegel wrote the *Logic*, then Hegel was brilliant' is of the form 'if p , then q ' and of the form 'if a is F , then a is G '.

Why think that recognition of logical form is essential to understanding? Well, it's commonly thought that our understanding of a thought or sentence consists in grasp either of its truth-conditions (Davidson 1967; Lepore & Ludwig 2002), or of the valid inferences to and from it (Brandom 1994), or both. Logical form contributes both to the truth-conditions of a thought and to its inferential role. So, whatever the right account is, recognizing logical form is essential.

Logical form is *general*. A given form makes a uniform contribution to every thought or sentence which has that form. For example, any thought of the form ' a is F ' will be true if and only if a is F ; any thought of the form ' $\neg p$ ' will be true if and only if p is false. This generality allows us to know things about thoughts and sentences that we don't understand. Consider the sentence 'Space is curved'. As long as I recognize the logical form of 'Space is curved', then – even if I don't understand the sentence – I know that 'Space is curved' is true if and only if space is curved (Higginbotham 1989).²⁶ I know, too, that 'Space is curved' entails that something is curved. We can draw inferences when we recognize a sentence as expressing a thought of a certain form – even if we don't know which thought.

This kind of understanding can come in stages, which it's natural to think of as levels of analysis. For example, take

(8) Space is curved and time is not real.

I may first recognize 8 as a conjunction:

8 is true if and only if 'Space is curved' is true and 'Time is not real' is true.

Next, I realize that the right hand conjunct is negated:

8 is true if and only if 'Space is curved' is true and 'Time is real' is false.

And so on. Dummett (1974) believed this process essential to the utility of logic: partial analysis reveals some licensed inferences, and further analysis lets us see further structure, licensing new inferences.

²⁶Higginbotham calls this the 'less demanding notion of meaning' (1989, 170; see also Higginbotham 1994, 102 and Armour-Garb & Woodbridge 2015, 158).

Can this process be extended to include non-logical concepts? In principle, it can. For example, suppose that I understand 'curved', but don't know what 'space' refers to: with this partial understanding, I might already grasp that 'Space is curved' entails that space is not flat. But whether these inferences are valid depends on whether 'curved' makes the same contribution to every sentence of which it is a part. It could be that some inferences from 'curved' are valid when we're talking about everyday objects and invalid when we're talking about space. In other words, it depends on whether non-logical concepts are general in the same sense as logic. So while, in principle, the process could be extended to non-logical concepts, such extensions are *risky* in a way that logic is not.

5.4.3 Pretentious nonsense: the general idea

We're now in a position to state a pretence account of nonsense. This is a pretence where something that is not meaningful is make-believedly meaningful. In this section I explain the general idea; in the next section, I work through two examples.

When we produce nonsense, we enter a pretence where a certain sentence expresses a thought, or where a certain mental event was a thought. The props are not thoughts. At the level of language, they are nonsensical sentences; at the level of thought, they are mental events: failed attempts at using certain conceptual capacities to think. Neither nonsensical sentences nor failed attempts at thinking stand in logical relations, but we can pretend they do.

The initial stipulation is that a given item is or expresses a thought. Not any old thought, though: a thought of a certain form. Given the stipulation that some bit of nonsense expresses a thought of a certain form, we can exercise our conceptual capacities just as we do with a partially-understood sentence. This allows us to reason with nonsense. When we make a nonsense-attribution, we're using the expressive resources of the pretence to point out the prop as nonsense.

The pretence account of nonsense satisfies both constraints. Engagement is satisfied because, as I will show, within a pretence we can both reason with nonsense and embed it in nonsense-attributions. Austerity is satisfied because we only pretend that nonsense expresses thoughts. Unlike the bank robbery game and others discussed by some pretence theorists (Armour-Garb & Woodbridge 2015, chs. 2 and 3), pretence does not operate here to allow a sentence to express a content other than the one it appears to express.

5.4.4 Examples of pretentious nonsense

Let's see how this works. I begin with an example of reasoning with nonsense and then give an example of a nonsense-attribution. Recall

- (3) Goodness is hexagonal.

An utterance of 3 doesn't say anything, but we may mistakenly suppose it does. In making this supposition, we enter a pretence defined by the following rules.

'Goodness is hexagonal' pretence

Props. The prop is the sentence 'Goodness is hexagonal'.

Initial Stipulations. 'Goodness is hexagonal' expresses a thought of the forms ' a is F ', ' a is hexagonal' and 'Goodness is F '.

Principle of Generation. If P is true, and if there is no set of make-believable true sentences $Q_1 \dots Q_n$ such that if $Q_1 \dots Q_n$ were true then P would not be true, then P is make-believable true.

Recursive Principle. If $P_1 \dots P_n$ are make-believable true and if $P_1 \dots P_n$ were true R would be true and there is no set of make-believe truths $Q_1 \dots Q_n$ such that if $Q_1 \dots Q_n$ were true then R would not be true, then R is make-believable true.

Before I discuss reasoning with 3, notice that we can simulate a truth-condition for 3 within the pretence. It's make-believable true that 3 expresses a thought of the form ' a is F '. We know that a meaningful sentence of this form is true if and only if a is F , so we incorporate this knowledge into the pretence using the Principle of Generation. Then, by the Recursive Principle, it's make-believable true that

- (9) 'Goodness is hexagonal' is true if and only if goodness is hexagonal.

Of course, 9 is nonsense, since its right-hand side is nonsense. So I don't propose to follow Higginbotham (1989, 156) in claiming that we can know purported truth-conditions like 9. Since 9 is nonsense, it is not true and so cannot be known. Or again, since 9 is nonsense, it does not express a thought, so, assuming the objects of knowledge are thoughts, it cannot be known. Rather, 9 expresses an illusion of understanding; the reasoning sketched explains its etiology.

Next, consider the apparent entailment from 3 to 4. First, note that it's make-believable true that 3 expresses a thought of the form ' a is hexagonal'. We know that

a meaningful sentence of the form '*a* is hexagonal' entails 4, so we incorporate this knowledge into the pretence by the Principle of Generation. Then it's make-believable, by the Recursive Principle, that 3 entails 4.

Let me comment on a few suspicious-looking aspects of this account. First, you might worry that we set out to explain the apparent validity of the inference from 3 to 4, not the apparent truth of the claim that 3 entails 4. In fact, there's no deep problem here. As stated, the Principle of Generation only allows for incorporating *truths* into the pretence. To solve this problem, we'd need to extend the Principle to allow for incorporating *rules of inference* as well. Just as what's make-believable true governs what's correct to assert within the pretence, what's make-believable valid governs what's correct to infer within the pretence. Given such an extension, we could incorporate a rule like 'From a thought of the form '*a* is hexagonal', infer that something is hexagonal', and use it to make-believable validate the inference from 3 to 4.

Second, you might worry about the Initial Stipulations about the form of 3. For example, the pretence contains the stipulation that 3 is of the form '*a* is hexagonal'. Are we entitled to this, if 3 doesn't express a thought? We are. Earlier, I argued against Diamond's view that words in nonsense sentences don't bear their ordinary meanings; rather, they do bear their meanings, but nonsense sentences don't express thoughts. So there's no reason to deny that 3 involves the ordinary phrase 'is hexagonal': a phrase which invites completion by some appropriate singular term. This obvious fact about 3 shapes the pretence into which we enter when we suppose that 3 is meaningful.

This fits with a plausible idea about nonsense. When we're under the illusion that 3 makes sense, we're not supposing merely that it expresses some thought or other – for example, that it might be a coded expression of a state secret. Rather, we're under the illusion that it expresses a thought which has components in common with 'Goodness is rare' and 'France is hexagonal'. Among other things, then, it must be of the form '*a* is hexagonal'. Why does 3 invite this illusion in particular? Because the illusions associated with a particular item of nonsense are generated and constrained by our ordinary ability to recognize patterns in speech and writing.²⁷

Next, let me address the use of nonsense in nonsense-attributions. The idea will be that, in order to make a nonsense-attribution, we have to enter into the pretence that the nonsense makes sense, and use the expressive resources of the

²⁷For further discussion of the role of pattern recognition in reasoning, see Besson (2019a).

pretence to point out 'from the inside' that it does not. This is what Evans called a 'game-to-reality shift' (1982, 369):

[S]omeone who utters such a sentence should be likened to someone who makes a move within a pretence in order to express the fact that it is a pretence. He is not like someone who tries to prevent a theatre audience from being too carried away by jumping up on the stage and saying: 'Look, these men are only actors, and there is no scaffold or buildings here—there are only props.' Rather, he is like someone who jumps up on the stage and says: 'Look, Suzanne and the thief over there are only characters in a play, and this scaffold and these buildings are just props.' The audience must be engaged, or be prepared to engage, in the make-believe, in order to understand what he is saying.

The analogue of a nonsense-attribution in the bank robbery game would be, 'all this cash is really just newspapers'.

I argued above that there are two readings of nonsense-attributions: a metalinguistic reading, on which the nonsense-attribution is about a string of symbols, and a non-metalinguistic reading. I want to suggest, now, that the non-metalinguistic reading is available only where there's a pretence in which the nonsense-attribution is make-believable: a pretence where the content of the that-clause make-believablely expresses a thought. Where there's no such pretence, the nonsense-attribution can only be read metalinguistically.

Where there is an appropriate pretence, some sentence or mental event will be used as a prop. It will be make-believablely true that the sentence in question expresses a thought, or that the mental event is a thought. Within the pretence, then, we gain an expressive resource for identifying the sentence or mental event in question: as that which expresses, or is, a particular thought. The nonsense-attribution exploits this expressive resource to identify the sentence or mental event in question and then states that – outside of the pretence – it does not express, or is not, a thought. When read in a non-metalinguistic way, then, the nonsense-attribution will be true if and only if the prop underlying the that-clause does not express, or is not, a thought.²⁸

Let's consider an example. Recall the nonsense-attribution

(5) It is nonsense to say that a picture is a fact.

²⁸While there's no room to develop this, I believe this treatment can be extended to deal with sentences like 'Wittgenstein believes that a picture is a fact'. The truth of this sentence will depend (loosely speaking) on Wittgenstein's participating in a pretence in which the embedded sentence is make-believablely true.

On the metalinguistic reading, 5 expresses the fact that the string ‘a picture is a fact’ is meaningless. But suppose there’s a pretence that ‘a picture is a fact’ expresses a thought. Within that pretence, 5 is make-believedly false.

Given that 5 is make-believedly false, a non-metalinguistic reading of 5 is available.²⁹ On this reading, 5 says that that which make-believedly expresses a thought – namely, the sentence ‘a picture is a fact’ – does not express a thought. Supposing that ‘a picture is a fact’ really is nonsense, 5 says something true. But it identifies the nonsense using expressive resources available only to one who has entered into the pretence that it makes sense.

I mentioned earlier that nonsense-attributions sound a bit odd. Now we can see why. Within the pretence, 5 is simply false, as ‘A picture is a fact’ is make-believedly meaningful. Outside of the pretence, we aren’t engaging with the nonsense, so 5 can only be read metalinguistically. And surely we don’t enter the pretence midway through 5, after ‘to say’. So what exactly is going on, semantically? Someone who utters 5 is speaking from within the pretence to point out that, in fact, it is a pretence. They are using expressive resources that are only available within the pretence to communicate something about the pretence itself. This kind of utterance is parasitic on the existence of nonsense; in an ideal language, it would have no place. But in a language where nonsense exists, it has real cognitive value.

5.4.5 Objections and replies

This concludes my account of nonsense. Of course, semantic pretence accounts face no shortage of objections (Richard 2000; 2013; Azzouni 2018; Kroon 2018; Woods 2018). In this section I discuss two of the most serious.

First, as Jody Azzouni has noted, inference ‘is invariably language-wide in scope. Sentences, nearly enough, from any area of discourse, may be employed to deduce results.’ This raises the problem of how ‘to understand inferences when they involve both pretence and non-pretence sentences from which non-pretence conclusions are drawn’ (2018, 700-701). For present purposes, the worry concerns reasoning that moves from partly nonsensical premises to non-nonsensical conclusions. Suppose we go from 10 and 11 to 12:

²⁹This non-uniform treatment of nonsense-attributions entails some non-compositionality. For example, while ‘It is nonsense to say that a picture is a fact’ involves a game-to-reality shift, ‘It is not nonsense to say that a picture is a fact’ is more naturally understood as make-believedly true. In fact, this is what we should expect. Compare an actor on stage who insists: ‘Suzanne and the thief over there are not just characters in a play!’

- (10) A picture is a fact.
- (11) If a picture is a fact, then some facts are hard to understand.
- (12) Some facts are hard to understand.

What's happening here? 10 is wholly nonsense. 11 is (let us suppose) also nonsense because of its antecedent.³⁰ But 12 is fine. The problem is that the reasoning from 10 and 11 to 12 seems to be either irrelevant or invalid. For if it takes place within a pretence, it can be, at best, make-believedly valid, establishing only the make-believe truth of 12, and is thus irrelevant to the genuine truth of 12. And if it takes place outside of a pretence, it is invalid, vitiating by the meaninglessness of 10 and 11.

There's no real problem here. We should cheerfully accept the first horn of the dilemma: the reasoning from 10 and 11 to 12 is valid only within the pretence that 10 expresses a thought. So it is only make-believedly valid, and it shows only the make-believe truth of 12. None of this undermines the fact (if it is one) that 12 expresses a truth. It only means that 10 and 11 are not a route to knowing that truth.³¹

Second, you might worry that participating in a pretence requires the intention to participate in a pretence (Azzouni 2018, 693-694). But those who are under the illusion that some item of nonsense makes sense have no such intention. So they cannot be participating in a pretence.

I don't think that participating in a pretence requires the intention to participate in a pretence. Rather, it requires intentions to proceed in accordance with certain suppositions – suppositions which may not be true, but need not be known to be false. In the present cases, these are suppositions to the effect that a certain sentence is meaningful or that a certain mental event is a thought. In some cases (as in 3 above), anyone engaged in such reasoning would know that the sentences are nonsense. In other cases (as in the *Tractatus*), the usual speaker doesn't believe that the sentences in question are nonsense. Nor do they positively believe that these rather strange sentences make sense. Rather, they take for granted that the sentences make sense.

³⁰If nonsense is contagious in this way, we might use Bochvar's logic B3 (also known as the Weak Kleene scheme) to represent the interaction of nonsense with sense. Indeed, Bochvar (1937) interprets the third truth-value as 'nonsense' or 'meaningless'. However, the question of the correct 'logic of nonsense' is subtle and beyond the scope of this paper. See Halldén (1949), Åqvist (1962) and Goddard & Routley (1973, ss. 5.3 and 5.4).

³¹This leaves intact the cases where nonsense is used in coming to know that it is nonsense – e.g. the Liar Paradox and Russell on existence. In such cases, the fact that a given item is nonsense is not known as the conclusion of an argument with nonsensical premises.

Where the sentence is nonsense, proceeding on such a supposition suffices for entering a pretence.

It might seem puzzling how an ordinary speaker could intend to follow the rather technical rules that govern what is correct to say within a pretence. After all, ordinary speakers need not possess concepts like *logical form* or *make-believedly true*. But this problem isn't specific to nonsense. It's an instance of the more general problem of how speakers are related to the rules that govern what is correct to say within a language, the rules we codify in a theory of meaning. These rules are formulated using concepts like *truth-condition*, which speakers need not possess. The solution, in my view, is to distinguish between the practical ability a speaker has and the theoretical representation of that ability (Evans 1981; Fricker 1982; Peacocke 1986; Wright 1986b; Davies 1989). The theoretical representation should specify the features to which a speaker is sensitive, but it need not do so in terms which a speaker would recognize. In the case of nonsense, the rules for a pretence specify the features to which speakers are sensitive when they participate in the pretence, but they need not do so in terms which speakers would recognize.

5.5 Conclusion. Is there a transparent level of sense?

I've argued that an account of nonsense must meet two constraints: the Engagement Constraint and the Austerity Constraint. While many existing accounts fail to meet one constraint or the other, I've proposed an account which meets both. Of course, I haven't shown that there are no other ways of accounting for nonsense. So, while I claim that pretence is what explains our engagement with nonsense, this claim is subject to other explanations that might be devised.

I'll end by setting out a hard question which should be of interest to anyone who finds the idea of nonsense compelling.

Philosophers have often thought that illusions of sense can be long-lasting and difficult to overcome: while it may only take a second's reflection to see through 'Goodness is hexagonal', it may take a lifetime to get past deep-rooted logico-metaphysical confusions (Moore 2003a, 185). This raises the prospect of *content skepticism*: skeptical doubt about whether we are really expressing thoughts. Can we ever be certain that we are having a thought, or may our attempts always turn out to be nonsense?³²

³²We should distinguish this from a global skepticism, about whether all of our sentences might

We could reject this skepticism if there were a level of thoughts where illusions of sense are impossible: a level where there's no distinction between sense and the appearance of it. Call this a *transparent level of sense*. Such a level of sense would be analogous to a level of perceptual content at which it's impossible to be mistaken: for example, while you can be mistaken about whether you're listening to a trumpet, or looking at a copy of *Naming and Necessity*, perhaps you can't be mistaken about whether you're hearing a brassy tone, or seeing an orangey-red hue. Of course, it's disputed whether there is such a level of perceptual content. And it's just as unclear whether there is a transparent level of sense (cf. Millikan 1984, 92).

Many early analytic philosophers thought there was. Frege and the early Wittgenstein, for example, thought that a logically perfect language would display its structure with total lucidity, such that nonsense could never appear to be sense. A statement like 'the Good is more identical than the Beautiful', expressed in such a language, would be an incoherent jumble of signs (*TLP* 4.003). Later philosophers, like Austin and Ryle, may have implicitly supposed that the domain of transparent sense was the domain of ordinary language. When we go to the store and ask where the toothbrushes are, or give somebody directions to the library, there is just no prospect that our utterances will turn out to have been nonsense.

But I'm not sure. Nobody ever managed to find the logically perfect language, and the line between ordinary and non-ordinary language has proven hard to draw. It seems clear, though, that the question of a transparent level of sense is closely bound up with the problem of how meaning can belong both to theory and experience. This is a problem for future research.

turn out to be nonsense. It might seem possible to hold that we can know whether we are having a thought, that there is a knowable condition for sense, while allowing that all our attempts might fail to meet it. This position is, however, fraught. The condition for having a thought would itself be the content of a thought; but then it isn't clear how we could know the condition. These problems of self-undermining are endemic to the philosophical terrain (Kripke 1982, 69-71; Heal 1989, ch. 6).

6 The value of thinking and the normativity of logic

Note. In Chapters 1 to 4, I argued that logic is constitutive of thinking. The present chapter is about how to parlay this claim into an account of the normativity of logic. Like Chapter 5, this one is written as a standalone paper. (It's based on my paper of the same title, forthcoming in *Philosophers' Imprint*; as a result, there is some summary of material from earlier chapters.) It defends a conditional claim, taking the constitutivity of logic as a supposition rather than as a result established earlier in the dissertation. In fact, the antecedent of the conditional is weaker than the particular view I've established; the main argument of this chapter would succeed given any member of a class of constitutive views which I describe in 6.1.

6.1 Introduction

Discussions of how logic relates to thinking tend to take one of two approaches. Some emphasize that logic is normative for thinking: it tells us how we ought to think, or what it is to think well. Others emphasize that logic is constitutive of thinking: it tells us what it is to think at all.¹ This paper is about how to bring these approaches together. In particular, the paper is about how to build an account of the normativity of logic around the claim that logic is constitutive of thinking.

Let me start with some background. First, the term 'thinking' is used in various ways. Sometimes it is used as a catch-all for activity with representational content – as, for example, by Descartes in the Second Meditation: 'Well, then, what am I? A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, wants, refuses, and also imagines and senses.' I am using the term in a more discriminating sense. 'Thinking' is not a term for the genus, representational activity, but for a particular species of representational activity. I take thinking to be the representational activity composed of acts of judging and inferring, acts whose contents are propositions or 'thoughts'. In a full account of thinking, we might also want to include acts such as entertaining a thought or reasoning under a supposition,

¹For the former approach, see Field (2009). For the latter approach, see Putnam (1992, 247).

but in this paper I leave these aside to focus on the core cases. I'm also assuming, for purposes of this paper, that judging is binary rather than allowing for varying degrees of confidence.²

Second, many discussions of the normativity of logic use 'logic' to refer to a theory of the sort produced by logicians, or to the discipline to which these theories belong.³ Typically, these theories are about a consequence relation that holds among sentences of a formal language. This is not what I will mean by 'logic'. I will mean a set of inference-rules that apply to thoughts in virtue of the way they are composed. For example, take Modus Ponens:

Modus Ponens From p and *if p , then q* , infer q .⁴

It is a nice question whether the consequence relation generated by the logical rules applicable to thoughts coincides with any consequence relation studied by logicians, such as classical or intuitionistic consequence, but since nothing in this paper turns on this issue, I take no stand on it here. Throughout the paper, I will use rules such as Modus Ponens as examples, but nothing turns on any particularities of these rules; if you're persuaded by counterexamples to Modus Ponens (McGee 1985; Kolodny & MacFarlane 2010), you can substitute a different example.

There are several further assumptions I'm going to make about logic so that I can focus on the question this paper is really about. I'm going to assume that we have a way of demarcating logical from non-logical rules. I'm going to assume, contrary to the arguments of logical pluralists (Beall and Restall 2005; Shapiro 2014), that there is a unique set of rules that govern thoughts – what is often referred to as the 'One True Logic'. The relation between pluralism and logical normativity is too complex to address here (Kouri Kessel & Shapiro 2017; Blake-Turner & Russell 2018; Steinberger 2019b). It's worth noting, though, that my arguments would also hold on the weaker view that, while there are multiple correct logics, there is a 'minimal kit' (Hale 2002, 299; Finn 2019b) of rules that hold in any correct logic.

Most importantly, I am going to assume that logic is constitutive of thinking. When I say that logic is 'constitutive' of thinking, I mean that representational

²In 6.2 and 6.5 I flag two places where relaxing the binary assumption might make a difference to the argument.

³For example, in arguing that logic isn't normative, Russell (2020) takes the thing that isn't normative to be a theory specifying a consequence relation on a language.

⁴It's natural to write down logical rules as imperatives, but this notational choice is not meant to suggest that the rule on its own has normative force. For this point, see Harman (1986, 5) and for some comments on how my view relates to Harman's, see 6.5 below. For present purposes, all we need is a way of specifying the class of transitions that the rule licenses.

activity must tend to conform to logical rules if it is to count as thinking at all. To be clear, this doesn't exclude the possibility that thinking is also constituted by some non-deductive – or even non-logical – inference-rules. It states a necessary condition for thinking, not a sufficient one. By way of initial motivation for this position, observe that there seems to be a limit on just how illogical someone's representational activity can be while they still count as thinking. As Jane Heal (1989, 89) puts it, '*completely* unsuccessful and chaotic thought, thought in which no shred of truth or rational connectedness is discernible, is an incoherent notion.' This limit on illogicality is just the flip-side of the constitutive role of logic.

In fact, there is not just one constitutive position but a family of positions, all of which hold that some conformity to logic is required for thinking, but which vary according to the kind and extent of conformity required. The strongest constitutive view, that thinking requires perfect conformity to logical rules, is held by Kimhi (2018) and has been attributed to Kant (Putnam 1992; Tolley 2006) and to Wittgenstein (Conant 1992). A weaker constitutive view, that thinking requires only some degree of conformity to logical rules, is held by Quine (1960), Davidson (1973; 1985) and Stich (1990). In this paper, I'm going to work with a view on which a subject's representational activity counts as thinking only if it manifests a disposition to conform to logical rules. Call this 'the dispositional-constitutive position'. Closely related positions have been attributed to Reid (Rysiew 2002), and defended by Heal (1994; 1999) and Wedgwood (2017). However, the main line of argument from sections 6.2 to 6.4 will also apply to the other constitutive views. I should note that the view of MacFarlane (2002) and Leech (2015) – that it is constitutive of thinking that it be assessable by logical norms – is not a constitutive view in my sense: as MacFarlane (2000, 54) makes clear, the view does not require any degree of conformity to logic.⁵

⁵Beyond this definitional point, there are good reasons not to think of the MacFarlane/Leech view as giving logic a constitutive role. We can see this by asking what makes it the case that some activity X is assessable by logical norms. Not everything is so assessable, after all: snow-shoveling is not. There are three possibilities. First, that what makes X assessable by logical norms is that it conforms to them to some degree. In this case, the MacFarlane/Leech view collapses into a properly constitutive one, but assessability is no longer what is constitutive of thinking – partial conformity is. Second, that what makes X assessable by logical norms is some non-logical feature. In this case, that non-logical feature is what is properly constitutive of thinking; assessability by logic is derivative. The third possibility is that it is a brute fact that X is assessable by logical norms. In this case, logic is properly constitutive of thinking, but at the cost of implausibility. Surely there are many non-normative differences between thinking and snow-shoveling that explain why the former is logically assessable and the latter is not. I think the second interpretation is likely the right one, in which case logic has no properly constitutive role on the MacFarlane/Leech view.

This paper does not aim to defend the dispositional-constitutive position. Rather, the paper is about whether and how an account of the normativity of logic can be built around logic's constitutive role. This is a complex issue. One question is whether the dispositional-constitutive position is on its own sufficient to account for logic's normativity, or, more broadly, whether it plays any part in an account of logic's normativity. Another question is whether the dispositional-constitutive position is even consistent with the normativity of logic.

A 'no' to the second question would, of course, entail a 'no' to the first. To start with the second question, then: the reason why you might worry that the dispositional-constitutive position is inconsistent with logical normativity is as follows. A standard is normative for the members of a given class only if it is possible for there to be a member of the class which fails to meet the standard.⁶ But if logic is constitutive of thinking, then there can't be thinking which fails to meet whatever standards logic sets, so those standards cannot be normative for thinking.

Or, to put the worry another way, if logic is constitutive of thinking, then logic doesn't tell us how we ought to think but rather what it is to think. It doesn't divide good thinking from bad thinking – it divides thinking from non-thinking. And the claim that logic divides good thinking from bad thinking (in a broad sense of 'good' and bad') seems to be a mere notational variant of the claim that logic is normative for thinking. So if logic is constitutive of thinking, then it is not normative for thinking.

However, this worry rests on misunderstanding the dispositional-constitutive position. That position, as I've stated it, does not entail that any time a subject violates a logical rule, the subject fails to think.⁷ What thinking requires is that the subject *tend* to conform to logical rules; mistakes are possible as long as the subject conforms to the requisite degree, where this conformity is explained by an underlying disposition to conform. So there can be thinking which fails to meet the standards logic sets. In other words, logic can both divide thinking from non-thinking and

⁶For detailed discussion see Lavin (2004). Leech has objected (2017, 366-7) that logical normativity requires neither freedom of choice of how to think nor the possibility of failure to accord with logical rules: 'a perfectly rational being would still be *right*'. I agree that normativity does not require free choice, but I do not think normativity can exist without some possibility for error. A perfectly rational being would still be right, but only because there could exist imperfectly rational beings who could be wrong (Nunez 2018, 1156 n. 10). Still, the strength of 'possibility' here is fairly weak: for a standard to be normative for a given class it need only be conceptually possible for there to be a member of the class which fails to meet the standard.

⁷In Lindeman's terms (2017, 235-6), it is Threshold Constitutivism, not Naïve Constitutivism.

divide good thinking from bad thinking. The constitutive role of logic is at least consistent with logical normativity.

This leaves us with the first question: whether the dispositional-constitutive position is on its own sufficient to account for logic's normativity, or, more broadly, whether it plays any part in an account of logic's normativity. The aim of the paper is to respond to this question. In other words, the paper is about how to get from the dispositional-constitutive position to the conclusion that logic is normative.

My answer will be that while the dispositional-constitutive position is not on its own sufficient to secure the normativity of logic, neither is it irrelevant to logic's normativity. Rather, the constitutive role of logic is essential to its normativity. The main points of the account I will develop can be compactly stated as follows:

1. Logical rules are constitutive of thinking.
2. Thinking is necessary for human flourishing.

In my view, it is because logical rules are constitutive of a good that those rules are normative.⁸

The paper goes as follows. In 6.2 I develop a natural line of thought about how to develop the constitutive position into an account of logical normativity by drawing on constitutivism in metaethics. In 6.3 I argue that, while this line of thought provides some insights, it is importantly incomplete, as it is unable to explain why we should think. I consider two attempts at rescuing the line of thought. The first, unsuccessful response is that it is self-defeating to ask why we ought to think. The second response is that we need to think. But this response secures normativity only if thinking has some connection to human flourishing. In 6.4, drawing on neo-Aristotelian theories of value, I argue that thinking is necessary for human flourishing. Logic is normative because it is constitutive of this good. In 6.5 I show that the resulting account deals nicely with problems that vex other accounts of logical normativity.

6.2 Logical self-constitution

In this section I sketch Christine Korsgaard's constitutivist account of the normativity of practical reason and then develop a parallel account of the normativity of logic.

⁸The two-part structure of my account is distinctive. Some, like Korsgaard (2009) and Nunez (2018), take the constitutive claim to be sufficient on its own for logical normativity. Others, like Wedgwood (2017, 207-8), take the constitutive claim to be (albeit correct) irrelevant to logical normativity.

6.2.1 The constitutivist account of practical reason

Korsgaard argues that the principles of practical reason are constitutive of action: they are built in to what it is to act. This, she argues, is the source of their normativity.

Take, for example, the principle of instrumental reason: that if you intend some end then you should intend the necessary means to the end. Korsgaard argues that this principle is not imposed from outside on agents: rather, it is constitutive of acting. For what distinguishes an action from a mere event is that an action is the result of the agent's intention, or 'willing an end'. And '[t]o will an end just is to will to cause or realize the end, hence to will to take the means to the end' (2008, 56). Acting on the principle of instrumental reason, then, is not one among various ways of acting; it is what it is to act.

You might worry that this means we can never will an end and fail to will the means: if we fail to will the means, then we must not have willed the end in the first place. It would follow that we can never violate the principle of instrumental reason. But Korsgaard argues that this is a mistake. It fails to appreciate the first-personal character of the principles of practical reason: they articulate what we commit ourselves to in acting.⁹ Thus, 'willing an end just is committing yourself to realizing the end' (2008, 57). As we can fall short of our commitments, we can sometimes will an end without willing the means. We cannot fall short to an arbitrary degree, however: at some point we are no longer acting at all.

Korsgaard gives similar arguments for more robust principles of practical reason, such as the Categorical Imperative. The upshot of these arguments is that the principles of practical reason are 'internal' or 'constitutive' standards: 'standards that a thing must meet in virtue of what it is' (2008, 112). As long as what you are doing is acting, your action is a good action insofar as it meets the internal standards of action. Performing bad actions, therefore, 'is not a different activity from performing good ones. *It is the same activity, badly done*' (2008, 113).

Korsgaard suggests that internal standards do not need the same kind of justification as 'external' ones. If being habitable is part of the internal standard for a house, then if you're going to build a house, there is no room for the question

⁹There's a difficult question about how Korsgaard's notion of *commitment* is related to the notion of a *disposition*. I don't want to claim that commitments are merely dispositions in disguise, but Korsgaard's view that there is a limit on how far we can fall short of our commitments while still having them brings the two notions close together. For further discussion of commitments, see Bilgrami (2012).

why you ought to build a habitable one. Similarly, if practical reason is the internal standard for action, then if you're going to act, there is no room for the question why you should act in accordance with practical reason.

Of course, the question immediately arises: why should we 'act', in this sense? Why should we 'will' an end, in the sense in which that requires willing the means, rather than just following our desires? Korsgaard's deepest answer to this question is that acting is how we constitute ourselves as unified agents. We are confronted with various sorts of temptations to act in different ways, and this is a sort of disunity. If I simply follow the temptations, my conduct can be attributed to those temptations, but not to me as a whole person. By conforming to the requirements of practical reason, I unify myself in the face of this play of temptations. 'For to will an end is not just to cause it, or even to allow an impulse in me to operate as its cause, but, so to speak, to consciously pick up the reins, and make myself the cause of the end.' If I never will an end, 'this means that I, considered as an agent, do not exist' (2008, 59-60).

6.2.2 The constitutivist account of logical normativity

Korsgaard suggests (2009, 67) that a parallel account can be given of the normativity of logic. The basic idea must be that logical rules are constitutive of thinking, just as principles of practical reason are constitutive of acting. In this section I sketch such an account. It will be helpful to begin with a line of thought that supports the constitutive position before developing an account of logical normativity.

Let's start with the claim that having a thought requires a 'setting' (Heal 1994; Stroud 1979). It doesn't make sense to suppose that someone could think a single thought as the entirety of their mental life. Nor does it make sense to suppose that someone could think a thought while being wholly insensitive to its relations to other thoughts. Having any thought requires grasping some of these connections: seeing the ways that other thoughts support it, or follow from it, or stand in tension with it. (It doesn't follow that, as inferentialists claim, there are 'canonical' relations the acceptance of which is a necessary condition on understanding.)

Now, it is impossible to grasp these connections between thoughts without being disposed to conform to some logical rules. These rules govern the connections between thoughts which make understanding possible. Some examples will help to make this clear, although I should emphasize that the general thesis can survive even if particular instances turn out to be problematic. First, it seems that it would be

impossible to grasp that one thought supports another without being disposed to follow Modus Ponens, which says, in intuitive terms, to believe q if you believe p and you take it that p supports q (Russell 1912, ch. VII). Or again, it seems that would be impossible to grasp that two thoughts are in tension without some aversion to believing contradictions – in other words, without a disposition not to believe p and *not-p*. Without some tendency to reason in these ways, it would be meaningless to take two thoughts to stand in relations of support or tension.

It follows that thinking requires dispositions to conform to some logical rules; in this sense, logical rules are constitutive of thinking.¹⁰ Now, it can't be the case that such dispositions preclude ever making a mistaken inference, or mistakenly rejecting a valid one. Rather, the dispositions will manifest in a tendency to conform to the rule, except where the disposition is overridden by other factors – for instance, inattention, tiredness or the complexity of the thoughts under consideration. But while the dispositions need not manifest in every case, they must exist if the subject is thinking at all.

In this way, the constitutivist will argue, logical rules are standards that thinking must meet in virtue of what it is – internal standards for thinking. If you're going to think, then there is no room for the further question of why you ought to think logically.

As before, however, the natural question is why we should think in the first place. The parallel with willing an end suggests the following response: in thinking, we constitute ourselves as unified subjects. We are confronted with a variety of impressions – temptations to take reality to be one way or another. This is a sort of disunity. If we passively followed these impressions, we might have various representations, but we would not be subjects – the loci of points of view on reality.¹¹ To be subjects, we have to actively commit ourselves to reality's being one way or another; we have to 'make up our minds' (Valaris 2017).¹²

¹⁰The existence of such dispositions might seem doubtful, given evidence that people tend to reason incorrectly in some situations (Wason 1968). In response, I would make two points. First, while the existence of these dispositions is an empirical question, the way this evidence bears on them is not straightforward. For example, in the Wason selection task it is unclear whether test subjects are engaged in conditional reasoning at all (Sperber, Cara & Girotto 1995, s. 2). Second, the dispositions posited by the constitutivist are general, and can fail to manifest in particular cases, as long as there is an explanation for the failure.

¹¹Compare Lewis's (1982) suggestion that we tolerate inconsistencies by fragmenting our total body of beliefs.

¹²I do not mean to suggest that perfect unity is possible for beings like us. One reason for this comes from cases like the Preface Paradox (Christensen 2004), where you believe p , q , r ... and

Let's call this the 'constitutivist account of logical normativity'.

6.3 A problem for constitutivism and two unsuccessful responses

The constitutivist account of logical normativity faces an immediate problem: it does not tell us why we ought to think. In this section I develop the problem and then consider two unsuccessful responses to it. The first response is a dialectical one, suggesting that the problem undermines itself, and the second response appeals to an innate need to think. I argue that both attempts are unsuccessful, but the second one points the way to a better approach.

6.3.1 The absence of value problem

We can see the problem by considering an objection made by David Enoch against constitutivist accounts of practical reason. In the words of Enoch's skeptic:

Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can't act without aiming at self-constitution, but why should I act? If your reasoning works, this just shows that I don't care about agency and action. I am perfectly happy being a shmagent – a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of shmagency) of self-constitution. I am perfectly happy performing shmactions – nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of self-constitution. (2006, 179)

This is what we might call, following Shamik Dasgupta (2018), an 'absence of value' problem. It may be that a certain concept – being unified, being an agent – will not apply to you unless you meet certain standards. But this fact is, for all that has been said so far, 'normatively inert' (Dasgupta 2018, 310). Why does it matter whether you are an agent?

Precisely the same problem will apply to the constitutivist account of logical normativity. Suppose it is true that your representational activity will not merit the

also that, since you are fallible, at least one of your beliefs $p, q, r \dots$ is false: an inconsistent set of beliefs. The unavoidability of such cases for a rational and fallible subject (Harman 1984, 109; 1986, 16) might make them less bad, but they remain a source of disunity. Of course, the Preface Paradox doesn't arise in a picture that uses degrees of belief only, but even in a degree picture there is little reason to think that any of us could achieve a totally coherent body of beliefs.

title of ‘thinking’ unless it tends to conform to logical rules. Why does it matter whether your activity merits this title? Perhaps if you fail to think, you fail to be a subject – but why does it matter to be a subject?

While Enoch puts the problem in the mouth of a skeptic, there is nothing particularly skeptical about it. The problem is better seen as a failure to explain where the normativity comes in. It may be that someone who does not think is not a subject at all, but this is consistent with there being nothing normatively significant about being a subject. More needs to be said in order to explain why failing to be a subject is a normatively significant failure.

We can see what is missing if we look at the problem from a different angle. Enoch rather dismissively refers to the failure to meet an internal standard as a failure to be ‘classified’ in a certain way.¹³ The point of his rhetoric is, I think, this: for any concept K which has an internal standard of the sort we have been discussing, there is, of course, another concept J with a different internal standard, such that something which fails to be a K may still be a J; indeed, there will even be a concept K*, such that something falls under K* if and only if it fails to meet the internal standard for K. For example, you count as ‘nonthinking’ if and only if you fall below the threshold for thinking. So, for the constitutivist to show that some standard is normative, it is not enough to identify some concept to which that standard is internal. The constitutivist has to show that this concept is distinguished; they have to give a reason why it matters to be thinking rather than nonthinking.

It seems to me, however, that, when we look at the problem this way, things are not as hopeless for the constitutivist as Enoch supposes. Is there really nothing we can say about why it matters to be thinking rather than nonthinking?

6.3.2 The dialectical response

Here is a first response to the problem. Someone who raises the objection discussed in the previous section is asking why they ought to think. In asking this question, however, the objector is inviting the constitutivist to give an argument in response. And in inviting argument, the objector is already committed to the practice of accepting claims on the basis of argument – i.e. to inference. So the objector is already committed to thinking rather than nonthinking.¹⁴

Why does the objector have to be committed in this way? Well, if they are not

¹³Cf. Railton’s worry (1997, 309) that constitutive arguments are ‘merely linguistic’.

¹⁴For a similar argument in the context of agency, see Silverstein (2015).

so committed, then it is wrong to describe them as making an ‘objection’, because this presupposes a commitment to thinking rather than nonthinking. And if they have not made an objection, then the constitutivist has nothing to worry about. Call this the ‘dialectical response’.

I don’t think we should be satisfied with this response. It depends on framing the absence of value problem as an objection pressed by an objector. This is what opens the door to arguing that the objector’s position is self-undermining. But we are not obliged to frame the problem in this way.

The constitutivist account was supposed to explain why logic is normative. In terms introduced by Dummett (1974), this is an ‘explanatory’ project rather than a ‘suasive’ one: the aim is not to persuade someone who denies that logic is normative *that* it is, but rather to explain to someone who accepts that logic is normative *why* it is.¹⁵ The absence of value problem suggests that the constitutivist has not fulfilled this explanatory task until they have shown why the concept of which logical rules are constitutive – the concept of thinking – is normatively significant. The problem, then, need not be framed as an objection pressed by an objector: we can see it more simply as a gap in the constitutivist’s own account.

6.3.3 The need to think

The second response to the absence of value problem is to argue that we simply have to engage in the activity of which the relevant rules are constitutive. Some of Korsgaard’s remarks suggest this approach (2009, 32): ‘the laws of practical reason govern our actions because if we don’t follow them we just aren’t acting, and acting is something that we must do.’ Similarly, the response to the question ‘why should I think?’ may be that you have to.

This claim can be understood in a couple of different ways. One reading would be that it is simply impossible not to think (i.e. there is no possible world where you are not thinking). But this is not the reading we should adopt. First, it seems straightforwardly false, since there are times when we are not thinking – for example, in deep sleep. Second, it is inconsistent with the claim that sometimes we fail to think in virtue of falling too far from the constitutive rules of thinking.

On a different reading, the idea is that the aim of thinking is inescapable. While we we don’t always succeed in thinking, we are always driven to do so. This reading

¹⁵Note that some moral constitutivists take their project to be a suasive one, with the aim of refuting a moral skeptic. I am assuming that this is the wrong approach when it comes to logic.

seems more promising. To fill it out further, I'm going to appropriate some work by Imogen Dickie on the mind's 'need to represent things outside itself'.¹⁶

Dickie characterizes a 'need' as a 'personal-level goal-representing state' (2015, 280) which, 'like an intention, can guide action, but which, unlike an intention, does not have propositional content' (100). To be a goal-representing state is, roughly, to represent some target and to guide us towards that target. Needs thus stand to intentions as perceptions (on non-conceptualist views) stand to beliefs.¹⁷ Dickie proposes, then, that the mind has a need to represent things outside itself. The target of the need is representing; the need guides us towards that target, motivating us to represent when we are not already doing so.

We can appropriate this account to respond to the absence of value problem on behalf of the constitutivist about logical normativity. The answer to the question why we should think is that we need to. What distinguishes the concept of thinking from the concept of nonthinking is that only the concept of thinking picks out the target of our need. As Dickie argues, if you have a need for x , and the need motivates you to take certain steps to attain x , and those steps reliably lead to attaining x , then those steps are 'strongly justified'. For example, if your need to eat leads you to take steps which reliably lead to eating, those steps are strongly (albeit not absolutely) justified. So, having proposed that we have a need to think, the constitutivist can argue that, since following logical rules is not only a reliable means to thinking but constitutive of it, following those rules is strongly justified. That following logical rules has some positive justificatory status, which not following those rules lacks, seems like enough to address the absence of value problem.

I agree that needs can play this justificatory role. But Dickie's characterization of what it is to be a need is incomplete. Dickie characterizes a need as a non-conceptual motivational state, but that is not a sufficient condition: a need also has to be a state whose fulfillment is good for the subject who has the need. We can put this point more carefully by drawing on Wiggins' analysis of needs. For Wiggins, I have a need for x (if and) only if 'it is necessary, things being as they actually are, that if I avoid being harmed then I have x ' (1997, 10). Being 'harmed' means falling below 'some however minimal level of flourishing that is actually attainable' (13, italics omitted). We have a need to eat not only because we have a non-conceptual

¹⁶I say 'appropriate' because Dickie's account involves a very different set of issues and commitments to those discussed here.

¹⁷So we can have a need for x without having the concept of x . For this idea of non-conceptual content, see Peacocke (1992, ch. 3).

motivation to eat, but because there is some minimal level of flourishing we cannot attain if we do not eat.

To fix ideas, let us say that a non-conceptual motivational state is a ‘drive’. Where a drive’s fulfillment is necessary for some minimal level of flourishing, the drive is a ‘need’; otherwise, it is a ‘mere drive’. Now, the relation between needs (in the proper sense) and flourishing is essential to the justificatory role that needs can play.¹⁸ Mere drives do no justificatory work. Consider, for example, someone who has a drive to do nothing but chop onions all day long. This state may reliably generate its own fulfillment, but all that chopping is not thereby justified. The reason is that the chopping is not necessary for the flourishing of the person with the drive.

So suppose that the constitutivist proposes that we have a non-conceptual motivational state whose target is thinking. The constitutivist must then clarify whether this state is a need or a mere drive. If it is a mere drive, then it cannot solve the absence of value problem. The fact that we are driven to think is just like the fact that someone is driven to chop onions all day: it is normatively inert.

It seems, then, that the constitutivist should claim that thinking is a need: that it is necessary for some level of human flourishing. In fact, I think this is the right way to go, and I will develop this suggestion in the next section. But it’s worth noting that if thinking is a good for us, then this is why the constitutive rules of thinking are normative. The feature of needs which does the justificatory work is not that they have motivational force but that their fulfillment is good for the subject who has them.

6.4 The value of thinking

Let’s take stock. We were attempting to build an account of the normativity of logic around the claim that logic is constitutive of thinking. The initial proposal was that logical rules are internal standards for thinking, and that thinking is how we unify ourselves as subjects. The problem was there are many different concepts with their own internal standards, such that it was unclear why it mattered if we fell short of the standards for the concept of thinking in particular: why should we think? The first response was that it is incoherent to ask why we should think, because

¹⁸To be clear, I think that the ‘need to represent’ is a need in the proper sense. But the consideration Dickie cites in support of the need’s existence – its ability to explain our representational behaviour (2015, 127) – show only that it is a drive. Showing that it is a need would require more normative considerations.

asking why already shows a commitment to thinking. The second response was that we need to think. I argued that the first response failed, while the second response could succeed only if thinking were necessary for human flourishing – more broadly, only if thinking had some value.

In this section, after briefly considering a range of claims we might make about the value of thinking, I'll argue that thinking is necessary for human flourishing.¹⁹ Before I start, however, let me comment on the shape of the account and how it relates to constitutivism about normativity. One attraction of constitutivism is the prospect of grounding normativity in the thinner notion of satisfying the internal standard of a concept. Relatedly, constitutivism offers the prospect of an account of normativity which a skeptic can only reject at the cost of literal inconsistency. By supplementing the constitutivist account with a claim about the value of thinking, we abandon both of these prospects. Why, then, should a constitutivist be interested?

At least in the logical case, I don't think these prospects are what is most compelling about constitutivism. What I'm after is an explanation of logical normativity, and what I find compelling in constitutivism is the idea that logic tells us what it is to think: that if we don't even tend to conform to logical rules, we are not thinking at all. This, it seems to me, is an idea we have independent reason to accept. The interest of the account I will develop is that it makes this idea central to (though not exhaustive of) an explanation of logical normativity.

6.4.1 Possible views of the value of thinking

Once we are willing to supplement the constitutivist account with a claim about the value of thinking, a range of options opens up. Different claims about the value of thinking lead to different views of the normativity of logic. In this section I briefly survey the options, from the most minimal to the more robust, before discussing my preferred option.

1. Thinking is instrumentally valuable as a means to some particular end, which some people have and others do not. On this view, logic would be normative, but only for those who shared this end, and only instrumentally.
2. Thinking is instrumentally valuable as a means to some particular end which everyone has. On this view, logic would be normative for everyone, but only

¹⁹For a different attempt to bring together constitutivism with neo-Aristotelianism, see Hacker-Wright (2012). See also Thompson (ms.) and Lott (2014).

instrumentally. This gives logic its universality, but – as Hilary Kornblith has noted – such a strategy incurs a ‘substantial burden of proof’:

any attempt to gain universal applicability by appeal to goals that all humans in fact have will almost certainly run afoul of the facts. Human beings are a very diverse lot; some of us are quite strange. It is hard to imagine making a plausible case for any particular goal or activity which is genuinely universally valued. (Kornblith 1993, 367; cf. Foot 2001, 44).

3. Thinking is instrumentally valuable as a means to every end, so that if you have any ends at all, you need to think. This is because thinking is necessary for deciding how best to pursue your ends. Kornblith suggests a similar view of epistemic norms, noting that on such a view, ‘they are derived from our desires in a way which removes any mystery surrounding them’, but are ‘universal in their applicability and not merely contingent upon having certain values’ (1993, 372).
4. Thinking is valuable for its own sake.²⁰ However, this is consistent with the possibility of human flourishing without thinking: in other words, thinking may be valuable for its own sake but not be something we need to do.
5. Thinking is valuable for its own sake because it is necessary for human flourishing. In other words, thinking is something we need to do.²¹

Option 1 is implausibly weak: it gives logical rules the same normative force as the rules of chess. Options 2, 3 and 4 are less weak, and I think they are live possibilities. However, I am going to develop Option 5. It seems plausible to me that thinking is as strongly related to human flourishing as this claim says; moreover, I don’t think there is any special reason to prefer theoretical austerity from the outset. But I should stress that what comes next is one way of developing the constitutivist position, not the only way.

Let me briefly comment on the way that Option 5 explains the normativity of logic. If we accept Option 5 together with the dispositional-constitutive position, we have the following two claims:

²⁰As several philosophers have argued, this is not the same as ‘intrinsic’ value: something can be valuable for its own sake in virtue of its intrinsic properties, or in virtue of its relational properties (Korsgaard 1983; Rabinowicz & Rønnow-Rasmussen 2000).

²¹Compare Moore (2003b, 128) and Schechter & Enoch (2006, 707).

1. Logical rules are constitutive of thinking.
2. Thinking is necessary for human flourishing.

In order to arrive at the conclusion that logic is normative, we need the following principle: if X is constitutive of Y, and Y is necessary for human flourishing, then X is normative.

By way of brief motivation for this principle, consider G.E.M. Anscombe's accounts of the authority of the law (1978) and the morality of promising (1969), both of which use a strategy parallel to the one I propose for logic. In both cases, there is something constituted by a rule: the existence of a legal order is constituted by the rule 'obey the law'; the institution of promising is constituted by the rule 'keep your promises'. And in both cases, the thing in question is necessary for human flourishing: the existence of a legal order ensures some degree of security from arbitrary violence; the institution of promising underwrites human cooperation. As Anscombe puts it, these things have the kind of necessity Aristotle defined as 'that without which some good will not be obtained'.²² This, Anscombe suggests, is why the rules in question are normative.

The principle that if X is constitutive of Y, and Y is necessary for human flourishing, then X is normative simply makes explicit Anscombe's explanatory strategy. If that strategy is a plausible one, then so is the principle.

6.4.2 Thinking and human flourishing

In this section I argue that thinking is necessary for human flourishing (Option 5 of the previous section). I don't have a proof of this claim, but will offer four considerations which support it. Following this, I address the absence of value problem.

First, as I noted above, Dickie argues (2015, 127) that the need to represent explains why we form beliefs in some situations but not others:

If you have plenty to think about, you are not hungry for food for thought, and are less likely to take up the opportunity to think about a thing that an attentional perceptual link provides. If you are hungry for food for thought, you will seize upon the opportunity provided by an attentional perceptual feed, sustaining the attentional link, and forming and maintaining a body of ⟨That is Φ ⟩ beliefs even if the object you are

²²Aristotle, *Metaphysics* V 1015a20. The same definition underlies Wiggins' account of needs, discussed above.

attending to is an unexciting specimen with which you would not bother in a situation where the need was being fulfilled in other ways.

I suggested that these considerations only support the existence of a drive – not a need in the proper sense. But now it is worth noting that the drive to think does not seem pathological in the way that a drive to chop onions all day does. So there is at least *prima facie* reason to take this to be a need in the proper sense.

Second, for most other human activities, it seems possible to imagine a scenario in which someone has a flourishing life without the activity. For example, eating is typical of human beings, but we can imagine someone who goes on a fast, perhaps even to their death, for some worthwhile end. And while prior eating might be a necessary condition for fasting, eating plays no role in justifying the fast. But thinking is different. The only way I can imagine someone having a flourishing life without thinking would be if they had intentionally renounced it – perhaps as a religious act.²³ And this just means that the person's nonthinking is only justified if it is itself the result of – that is, justified by – thinking. So some thinking remains necessary.

It's worth emphasizing, too, just how deeply a life without thinking would differ from our own. As suggested in Option 3 in the previous section, thinking is necessary for deciding how best to pursue your ends. So it's not clear that someone who had renounced thinking could exercise agency. And it seems doubtful that we should count as flourishing a life that involved no exercise of agency.

Third, we might consider the capacities that are typical of human beings. One capacity that does seem typical is the capacity for rational activity. As Philippa Foot writes (2001, 56),

there is this great difference between human beings and even the most intelligent of animals. Human beings not only have the power to reason about all sorts of things in a speculative way, but also the power to *see grounds* for acting in one way rather than another; and if told that they should do one thing rather than another, they can ask why they should.

A more traditional way to put this point is to say that human beings are rational animals, or thinking animals. Now, in the classical neo-Aristotelian framework, a

²³A referee suggests that someone might renounce thinking as the result of a non-conceptual experience, such as an epiphany or revelation. If this were possible, it would weaken the force of the present consideration, perhaps motivating a retreat to a weaker claim (Option 1, 2, 3 or 4). But I think the possibility is somewhat tenuous: it might be more fitting to say, in such a case, that the person is not flourishing.

member of a given life-form cannot flourish without doing the things that are typical for members of that life-form (Foot 2001, ch. 2).²⁴ So if thinking is typical for the human life-form, then thinking is necessary for human flourishing.

Similar ideas are implicit in some constitutivist writing. Tyke Nunez (2018), for example, argues that logic is normative because logical rules specify the proper exercise of our capacities: ‘every exercise of the faculty ought to accord with its laws’ (2018, 1162). On its own, Nunez’ claim is not enough to secure normativity, as we can ask why we ought to exercise our capacities properly rather than improperly. Nunez’ claim needs to be supplemented with the claim that the capacity to think is characteristic of human beings, such that the proper exercise of this capacity is necessary for human flourishing.

Fourth, recall what I referred to as Korsgaard’s ‘deepest answer’ to the question of why we should act: that acting is how we constitute ourselves as unified agents. Along similar lines, I suggested that thinking is how we constitute ourselves as unified subjects. On their own, I argued, these ideas fail to secure normativity. But now we can see these ideas in a different, more Aristotelian light. We are living things of a particular sort – human beings. As living things, we are organized in a teleological way: we need to maintain our unity in the face of a tendency to disunity (Tenenbaum 2011; Moosavi 2019). As a general rule, then, a necessary condition for a living thing to flourish is that it maintain its unity.

But different sorts of living things are unified in different ways. Every animal is unified by maintaining the distinctness of its body from its surroundings. But as human beings, we are unified in a further, particularly self-conscious way – as agents and, more importantly for present purposes, as subjects. Given that thinking is how human beings constitute ourselves as unified subjects, and that such unity is a necessary condition for a living thing to flourish, it follows that thinking is necessary for human flourishing.

Finally, let me tie this back to the absence of value problem. One way of looking at that problem is that, even if there is a concept *K* to which some standard is internal, it’s not clear why it matters to be a *K* rather than something else. In other words, the constitutivist has to show that the concept they care about is

²⁴We might worry about the consequences of this Aristotelian claim for human beings who are, in various ways, unable to do what is ‘typical for the life-form’. In my view, the right response to this worry is to place less emphasis on the human essence and more emphasis on the idea that flourishing consists in the full exercise of the capacities which one actually has. (Compare Wiggins’ emphasis on the degree of flourishing that is ‘actually attainable’.)

distinguished in some normatively significant way. We're now in a position to address this problem. What distinguishes thinking from nonthinking is the role that thinking plays in the life of human beings. It is simply part of being a human being that thinking is necessary for your flourishing.

Of course, you might ask: why does it matter to be a human being, rather than a human being*, where the latter is like a human being, but without a need to think? Perhaps a human being* has a need to engage in some different representational activity, thinking*, which is somewhat like thinking, but not constituted by logical rules. Moreover, perhaps thinking* is better than thinking, so that we might be better off trying to be human beings* rather than human beings.

We should go slowly here. First, it's not clear what it would mean for thinking* to be better than thinking. Better how? To make sense of this claim, we would need some idea of a standard, applicable to both thinking and thinking*, which thinking* comes closer to meeting than thinking. And it is at least difficult to see what that standard would be.²⁵ (Consider the parallel question of whether thinking is better than perceiving.)

Second, it's not clear how to understand the idea that we might be better off trying to be human beings* rather than human beings. In particular, it's hard to distinguish this from the question whether it would be better if human beings were replaced by human beings*.²⁶ For it's not clear in what sense the resultant beings would be *us*.

The right response to the absence of value problem, then, is that there are limits on our ability to live by alternative concepts: it is not the case that for any concept *K*, we can invent a concept *K** which is an intelligible alternative for beings like *us*. There are some concepts – we might think of them as ‘bedrock concepts’ (Chalmers 2011b, s. 8) – for which we have no alternatives. I have suggested that the concept of thinking and the concept of a human being are bedrock in this sense. We are human beings, with a need to think: these facts are not up to us.

For these reasons, we should accept that thinking is necessary for human flourishing. Supposing that logic is constitutive of thinking, it follows that logic is normative.

²⁵Indeed, it is unclear what thinking* is supposed to be. For thinking is not just a name for representational activity which meets certain constraints: it is representational activity whose contents are thoughts. So we can ask what kind of content thinking* is supposed to have, if not thoughts.

²⁶For discussion of this question, see Williams (2006) and Moore (2018, ch. 17).

6.5 Vexing issues about logical normativity

That concludes the main part of my case for the proposed account of logical normativity. In this section, I argue that this account deals nicely with some vexing issues that arise in the literature on logical normativity. This will help to clarify the kind of normativity that logic has in my account and the role of this normativity in an explanation of how logic relates to thinking.

The issues I have in mind are about various implausible consequences that follow from certain formulations of the claim that logical rules are normative. For example, consider the Simple Formulation:

Simple Formulation If p entails q , then if you believe p , you ought to believe q .²⁷

This has a welter of implausible consequences: first, it seems to entail that if you believe p , you ought to believe *if p , then p* , and then you ought to believe *if p , then, if p , then p* and so on. This seems a waste of cognitive resources (Harman 1986, 12).²⁸ Second, it seems to entail that if you believe p and *if p , then q* , then you ought to believe q – even if q is false (Harman 1984; MacFarlane 2004).²⁹ And surely we ought not to believe things that are false. Third, given that p entails p , and there is nothing in the Simple Formulation to exclude the case where $q = p$, it seems to entail that if you believe p , you ought to believe p . This seems like an objectionable kind of bootstrapping: believing something doesn't on its own give you a reason to believe it (Broome 1999).

Now, the claim that logical rules are normative need not be committed to the Simple Formulation. But the underlying concern is that any formulation of the claim that logical rules are normative will entail similar consequences, or else be too weak to be interesting. Defenders of the normativity of logic have tended to respond by developing 'bridge principles', weakenings of the Simple Formulation that

²⁷A fully general formulation of this thesis would have to make provision for multiple-premise entailments. However, this detail doesn't matter for our purposes, as the problems arise even in the single-premise case. Also, if we conceived of believing in terms of degrees of confidence, we would need to reformulate the thesis to impose a constraint on the degrees of belief in p and q : for discussion, see Field (2009) and Milne (2009).

²⁸In fact, depending on what a belief is, the 'cluttering' beliefs might not count as additional beliefs. (For example, they don't narrow the set of possible worlds in which the subject's beliefs are true.) But for purposes of argument I will waive this point. Harman (1986, 14) responds to a related point by specifying that the objection applies to explicit rather than implicit beliefs.

²⁹Parallel arguments are often given against the claim that meaning is normative: see Hattiangadi (2006) and Glüer & Wikforss (2009).

avoid the implausible consequences. Here, I begin with some general comments on how these issues appear in my account. Next, I respond to each of the implausible consequences mentioned above. Finally, I compare my treatment of these issues with recent accounts centred on bridge principles, showing that, in my account, while a bridge principle holds, it is not fundamental in explaining the relation of logic to thinking.

6.5.1 Three questions about logical rules

To begin the response, we need to distinguish three questions about logical rules and their corresponding answers. The first question is: what are the logical rules constitutive of thinking? For present purposes, I will take Conditional Proof and Modus Ponens to be logical rules constitutive of thinking. I formulate them as follows:

Conditional Proof If by assuming p you can deduce q , infer *if p , then q* .

Modus Ponens From p and *if p , then q* , infer q .

All rules define a standard of correctness, specifying some acts as correct in light of the rule (Broome 2014, 24). It would be equivalent, then, to formulate the rules as follows:

Conditional Proof If by assuming p you can deduce q , it is correct to infer *if p , then q* .

Modus Ponens From p and *if p , then q* , it is correct to infer q .

As I mentioned earlier in the paper, the choice of how to present logical rules is a notational one; it doesn't yet answer any philosophical question about the normative force of the rules.

The second question is: what relation must a subject have to the rule in order to think? So far, I've said simply that a subject must have a disposition to conform to the rule, a disposition which is manifested in a subject's tendency to conform to the rule. But now I need to say a bit more about what this amounts to. Claims about dispositions play an explanatory role: when X has a disposition to do Y , and then X does Y , we can explain why X did Y by appeal to its disposition. But where the bearer of the disposition is a subject, we also have to say what it is first-personally for the subject to have, and to exercise, the disposition.

Borrowing from Peacocke (1987), I am going to say that when a subject has a disposition to conform to some logical rule, the subject finds the transitions that the rule specifies as correct ‘primitively compelling’.³⁰ That is, they are compelling, and the subject need not have any further idea of why they are compelling. To be clear, this means that the subject is tempted by particular inferences which evidently fall under the rule. It does not mean that the subject is tempted to accept a general representation of the rule.³¹ The subject can act on the resulting compulsion by making the transition, but there is no guarantee that they will do so: the question may never arise, or the compulsion may be overridden by competing factors. These factors might include inattention, tiredness, the complexity of the thoughts involved (making it unobvious that the transition falls under the rule) or other reasons not to make the transition.³² Unless the subject finds the transitions compelling, however, they are not thinking.

The third question is: why ought we to think? And here my claim is that thinking is necessary for human flourishing. By contrast, there are other rules such that we must have a similar relation to them in order to do a certain activity, but there is no general reason why we ought to do that activity. For example, we must have a similar relation to the rules of chess in order to count as playing chess, but there is no general reason why we ought to play chess.

Now, I think that full answers to these three questions would be an exhaustive account of the sense in which Conditional Proof and Modus Ponens are normative. But at no point in answering these questions are we committed to the Simple Formulation. The Simple Formulation is neither an answer to the first question (what the logical rules are), nor to the second question (how a subject must relate to logical rules in order to think), nor to the third (why the subject ought to think). Nor does it follow from the answers to these questions taken jointly.

6.5.2 The implausible consequences of the Simple Formulation

Let me now consider the implausible consequences that follow from the Simple Formulation. The first was that if you believe p , you ought to believe *if p , then p* and so on. The second was that if you believe p and *if p , then q* , then you ought to

³⁰Harman discusses a related notion of ‘psychological immediacy’ in Appendix A of his (1986).

³¹Most people would find it hard to resist the following reasoning: *Either I left my keys at home or in the car. They’re not in the car. So they must be at home.* But they need not find a representation of Disjunction Elimination intuitive.

³²There is a helpful discussion of the various possibilities here in Moore (2003b, 48-49).

believe q – even if q is false. The third was that if you believe p , you ought to believe p . Let me start with the first two cases, as the third raises some additional issues.

Here is what I want to say about the inferences in the first two cases. First, both inferences are correct in light of logical rules which we are supposing to be constitutive of thinking: the first in light of Conditional Proof, the second in light of Modus Ponens.

Second, this means that anyone who thinks has a disposition to conform to these rules. As I suggested above, this means that anyone who thinks finds the transitions that the rules specify as correct primitively compelling. But it does not follow that anyone who thinks must draw these inferences: the question may never arise, or the compulsion may be overridden by competing factors. In the case of the move from p to *if p , then p* , it seems plausible that the question will never arise. In the case of the move from p and *if p , then q* to q , where q is false, the fact that q is false is a competing factor that can, and should, override the compulsion to draw the inference.³³

Third, however, if the subject does not have the disposition to conform to these rules – i.e. if the question does arise, and there are no competing considerations, but the subject does not find the transitions compelling – then the subject is not thinking.

Consider a parallel case. Suppose that you believe *x is an $F G$* (for example, ‘This is a tall tree’). This entails – not logically, but bracket this for now – *x is a G* (‘This is a tree’). By analogy with the Simple Formulation, it might be proposed that if you believe *x is an $F G$* , you ought to believe *x is a G* . The same problems will arise: first, that this is a waste of cognitive resources, and second, what if *x is a G* is false?

Here is what I would say instead. First, the inference from *x is an $F G$* to *x is a G* is correct in light of the rule for thoughts of this structure.³⁴ Second, anyone who

³³An alternative solution to this problem is to weaken our characterization of the dispositions required for grasping logical rules. Rather than dispositions to *infer* a conclusion on the basis of *judging* the premises, Murzi & Steinberger (2013) propose dispositions to *consider* a conclusion on the basis of *entertaining or supposing* the premises. I am sympathetic to this solution, but accepting it would require discussion of the role of entertaining and suppositional reasoning in thinking, which I have no room to do here.

³⁴For discussion of such ‘structurally valid’ inferences, see Evans (1976) and Balcerak Jackson (2007). To give a full account of these inferences we would need to distinguish bad cases like ‘ x is a rubber duck, so x is a duck’. I’m going to assume that there is some way of distinguishing these cases, as it seems plausible that in good cases the inference in question is closely tied to understanding (Balcerak Jackson 2009).

understands *x is an F G* must have a disposition to conform to this rule; they must find the transition to *x is a G* primitively compelling. However, this does not mean they must draw this inference: the question may never arise, or the compulsion may be overridden. But, third, if they do not have this disposition at all, then they simply do not understand *x is an F G*. The fact that *x is a G* is false is a good reason to override the disposition, but it is not a good reason for denying the existence of the disposition altogether. (The parallel is only partial because the rule in question is not a logical one: you can fail to understand the relevant structure in *x is an F G* while still thinking. In general, this is the difference between the normativity of logic and the normativity, if there is any, of non-logical elements of content.)

Finally, the third implausible consequence of the Simple Formulation was that if you believe *p*, you ought to believe *p*. This raises somewhat different issues from the previous two cases because it doesn't depend on the rules for any logical connective. Rather, it depends only on Reflexivity, which is a structural rule – that is, an inference rule that is not about any logical connective.

Reflexivity From *p*, infer *p*.

This rule guarantees that *p* entails *p*, and then the Simple Formulation tells us that if you believe *p*, you ought to believe *p*.

In the discussion so far, I've implicitly taken 'logical rules' to refer only to operational rules – rules for logical connectives – and haven't said anything about the role of structural rules, or how thinkers must relate to such rules. Structural rules may raise different issues. While it seems plausible that we have dispositions to infer in accordance with Conditional Proof and Modus Ponens, it's less clear what it would mean to have a disposition to infer in accordance with, say, Transitivity. On the other hand, it's hard to draw a clear distinction between operational and structural rules, as operational rules can be seen as reflecting structural rules (Došen 1989), and operational rules can be seen as containing information about structural rules (Dicher 2016). These facts favour a uniform treatment of operational and structural rules.

Without taking a conclusive position on how other structural rules should be treated, it does seem plausible to me that any thinker has to have a disposition to conform to Reflexivity. If you judge *p* but do not find it primitively compelling to judge *p*, then there is reason to doubt that you are judging at all. The unintelligibility of judging *p* and refusing to judge *p* is even stronger than the unintelligibility of

judging *x is an F G* and refusing to judge *x is a G*. So I think we should take Reflexivity to be constitutive of thinking, and this means that we need a response to the bootstrapping problem raised above.

At this point my response is fairly similar to my response to the first two problems. First, the inference from *p* to *p* is correct in light of Reflexivity. Second, anyone who understands *p* must find this transition (really a degenerate case of ‘transition’) primitively compelling. Third, if someone does not find this transition primitively compelling, they are not thinking. None of this means that if you believe something, you ought to believe it.

6.5.3 Bridge principles

In this section I compare my approach to logical normativity with recent approaches centred on developing ‘bridge principles’.

A general strategy common to recent approaches to logical normativity is to argue that the normativity of logic consists in the holding of a ‘bridge principle’, which is a weakened version of the Simple Formulation above, of the following form:³⁵

Bridge Principle If *p* entails *q*, [normative statement about attitudes to *p* and *q*].

Bridge principles are developed so as to avoid the implausible consequences of the Simple Formulation. For example, a principle which avoids all three problems might be:

If *p* entails *q*, you ought not to believe *p* while disbelieving *q*, unless *q* is false.

As this principle does not enjoin you to believe anything, it does not enjoin cluttering your mind with useless consequences, nor believing the things you happen to already believe. Nor does the rule prohibit disbelieving a consequence of your beliefs if the consequence is false.

I’d like to make three remarks by way of comparing my account to the bridge principle strategy. This comparison will shed light on my own account, provide additional motivation for a particular class of bridge principles, and also raise a question about the larger explanatory role of bridge principles. First, I will show that

³⁵This literature begins with MacFarlane (2004); for more discussion see Field (2009), Broome (2013) and Steinberger (2019a). A fully general bridge principle would, of course, allow for multiple premises. It is also possible that more than one bridge principle is required, in order to capture different varieties of logical normativity (Steinberger 2019c), but I will ignore this detail here.

it is possible to generate a bridge principle from my account. Second, however, the bridge principle is open-ended, rather than attempting to specify when we should, or should not, draw a valid inference, and I will offer some principled reasons to think this is the right approach. Third, I will contrast the explanatory role of bridge principles on my account from the role they have in many discussions: it is often suggested that bridge principles bridge a gap between logic and thinking, but on my account this is not so.

First, then, it is possible to state a bridge principle in my account. To do this, we have to make a modification to the form above. The antecedent of the form above is ‘If p entails q ’, but in my account the relation that thinkers must be responsive to is not entailment in general, but entailment in light of particular logical rules. Let us say that p ‘directly entails’ q if and only if the transition from p to q is correct in light of a single application of a logical rule.³⁶ (For example, p directly entails *if p , then p* , but does not directly entail *if p , then if p , then p* .³⁷) The antecedent of my bridge principle must invoke direct entailment rather than entailment.

What about the consequent of the bridge principle? Initially, it seems like all we have is:

If p directly entails q , then if you believe p you must find the transition to q primitively compelling.

If we wanted to build in more detail, we could say:

If p directly entails q , then if you believe p , and if the question arises whether q , you have a reason to believe q , unless other factors override this.

But we have to be careful about ‘reason’ here. This ‘reason’ is just an articulation of the primitive compulsion a subject feels in virtue of having a disposition to conform to a logical rule. It is not a justifying reason (an ‘other things being equal it would be good if...’ reason), because a subject could have a similar reason in virtue of being disposed to conform to the rule for ‘tonk’ (Prior 1960). Rather, it is an explanatory reason – a motivational state which could explain the subject’s action (Parfit 1997; Alvarez 2017). And this means that we do not yet have a genuine bridge principle: a bridge principle has a normative statement as its consequent, but a claim about

³⁶We can define entailment as the transitive closure of direct entailment. p entails q if and only if for some set $\{p_1, p_2, \dots, p_n\}$, p directly entails p_1 , p_1 directly entails p_2 ... and p_n directly entails q .

³⁷Compare Field’s (2009) notion of ‘obvious’ entailment and the diachronic norms discussed by Hlobil (2015).

explanatory reasons is not a normative statement. Similar remarks apply to ‘must’ in the previous formulation of the principle.

However, in my account thinking does have genuine normative status, and a failure to find the right transitions primitively compelling means a failure to think. In other words, while the principle above simply describes what it is to have a disposition to conform to logical rules, the value of thinking means that we have a reason – indeed, a need – to have such a disposition. This, I think, allows us to say that if p directly entails q , someone who believes p does have a justifying (as well as explanatory) reason to believe q . If they refuse to believe q , this raises doubts about whether they are thinking. So we have:

If p directly entails q , then if you believe p , and if the question arises whether q , you have a reason to believe q , unless other factors override this.

This (fairly weak) bridge principle is true in my account.

To be clear, this bridge principle is not particularly novel. It is a reasons-based, rather than ought-based principle, like that of Sainsbury (2002). This puts it in tension with the ‘Strictness Test’ that MacFarlane (2004) draws from Broome (1999): essentially that *pro tanto* reasons are too weak to properly capture logical normativity. The intuitive worry is that reasons-based bridge principles make it too easy for other considerations to override logical reasons. In my view, some of the force of this objection comes from conflating explanatory and justifying reasons. The explanatory reasons generated by logical dispositions may, indeed, be hard to override: if you believe p , and p directly entails q , you can’t choose not to believe q merely because you are offered some money. But the justifying reasons generated by logic are *pro tanto*, as there may be good reason, all things considered, to try to bring it about that you disbelieve q , even if it follows from some p which you accept.

This leads to the second remark I want to make. The bridge principle in my account is open-ended, containing a reference to overriding considerations, rather than attempting to specify when we should, and when we should not, draw a valid inference. This is a feature it shares with other reasons-based bridge principles (MacFarlane 2004). This contrasts with some other investigations of the normativity of logic, which aim at finding a plausible bridge principle which avoids implausible consequences while still providing some reasonably strong normativity. The way these accounts avoid implausible consequences is by building a bridge principle explicitly to avoid them – for example, building in an exception for cases where the conclusion

of the entailment is false. The resulting bridge principle satisfies the Strictness Test, specifying when we should, and should not, draw a valid inference. My account generates some novel reasons of principle to think this is the wrong strategy.

To show this, I want to develop an analogy. Consider the normativity of promising. I think – as did Anscombe (1969) – that an account of the normativity of promising should be composed of two parts. First, there is the rule of promising – the rule you have to tend to follow in order to count as promising. Plausibly, this rule is simply ‘keep your promises, unless released from doing so by the promisee’. Second, however, there is an account of why promising is something we should go in for. Here there are a range of options: virtue theories, contract theories, consequentialist theories and so on (Habib 2018).

Now, what about a promise to commit murder? Such a promise should not be kept: so should we say that the rule of promising is really ‘keep your promises, unless released from doing so by the promisee, or unless the promise is to commit murder’? In my view, we should not. Having a tendency to follow the rule of promising does not require that you follow the rule in every case; it is consistent with this tendency that you sometimes override it when there are good reasons to do so. If we built all the exceptions there are into the rule of promising, what we would end up with would not be an account of the normativity proprietary to promising; it would be an account of general morality contained as exceptions to promise-keeping.

Similarly, I do not think we should incorporate the rule against believing falsehoods, or other good epistemic rules, into an account of the normativity of logic. The result would no longer be an account of the normativity proprietary to logic; it would be an account of good epistemic conduct. Of course, an account of the normativity of logic will be part of an account of good epistemic conduct – this much is reflected in the open-endedness of my bridge principle – but that does not mean that we should be able to read off the latter from the former, any more than we should be able to read off the wrongness of murder from our account of the normativity of promising. If this is right, then there are reasons of principle for rejecting the Strictness Test as a constraint on bridge principles.

Finally, I’d like to end by spelling out a deep difference between my account and a view of logic and thinking that motivates much work on bridge principles: namely, that these principles bridge a gap between logic and thinking. The idea that there is such a gap derives from Gilbert Harman (1986, 6). Steinberger explains the idea as follows (2019a, 307; citation omitted):

The traditional conception whereby logic occupies a normative role in our cognitive economy rests upon the mistake of conflating (or at least running too closely together) principles of deductive logic with what Harman calls ‘a theory of reasoning’. Yet the two enterprises—formulating a deductive logic and formulating a normative theory of reasoning—are fundamentally different according to Harman. A theory of reasoning is a theory of how ordinary agents should go about managing their beliefs. ... In short, Harman’s explanation of our intuitions to the effect that logic must have a normative role to play in reasoning is that we conflate deductive logic and theories of reasoning. Little wonder, then, that we take there to be an intimate relation between logic and norms of belief: the relation is simply that of identity! However, once we are disabused of this confusion, Harman maintains, we are left with ‘a gap’. The question is whether that gap separating logic and norms of reasoning can be bridged.

Bridge principles are then conceived as a way to bridge the gap. This is reflected in the form of the principles: the antecedent is a claim about thoughts or sentences, saying nothing about thinking, while the consequent is a claim about thinking. Given that the claim about thinking is a normative one, the suggestion is that what bridges the gap between logic and thinking is the normativity of logic.³⁸ If normativity is what bridges the gap between logic and thinking, and the correct bridge principle specifies that normativity, then the correct bridge principle is not just true, but also fundamental in an explanation of how logic relates to thinking.

The account I’ve developed in this paper has a very different shape. On my account, a bridge principle is true, but it is derivative, rather than fundamental, in an explanation of how logic relates to thinking. Of course, this point isn’t meant as an objection to Harman, or to the bridge principles literature: that would require defending the constitutive position, which I haven’t done here. But it is still worth spelling out the difference in approach, because it helps us to clarify the theoretical ambitions of an account of logical normativity.

On my own account, the correct bridge principle does not bridge a gap between logic and thinking. If there is such a gap, then what bridges it is that logic is constitutive of thinking – i.e. that thinkers have to tend to conform to logical rules in order to think. But given the way I defined ‘logic’ as a set of inference-rules operating on thoughts, it might be better to say that, on my own account, there is

³⁸This suggestion is not essential to the study of bridge principles: it would be possible to frame them as attempting to specify the way in which logic is normative, but not attempting to make any claims about what is explanatorily fundamental. However, as is clear from MacFarlane (2004) and Steinberger (2019a), the gap idea does motivate much of this research.

no gap at all between logic and thinking. As a result, it would be possible to give an explanation of how logic relates to thinking, and of why logic is normative for thinking, without any reference to bridge principles.

This possibility should force us to clarify what we are after in studying bridge principles. We might simply be interested in finding some truth about the normative force of logic. In that case, seeking a satisfactory bridge principle may be a good strategy. But – as suggested by talk of a ‘gap’ between logic and thinking – we might also be interested in giving a fundamental explanation of how logic relates to thinking. In that case, we should not take for granted that bridge principles are the place to start.

Bibliography

- Åqvist, L. (1962). Reflections on the logic of nonsense. *Theoria* 28(2), 138–157.
- Alvarez, M. (2017). Reasons for action: Justification, motivation, explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 ed.). Metaphysics Research Lab, Stanford University.
- Anderson, A. R. and N. D. Belnap (1975). *Entailment: The Logic of Relevance and Necessity, Vol. I*. Princeton University Press.
- Anscombe, G. E. M. (1969). On promising and its justice, and whether it needs be respected *in foro interno*. *Critica* 3(7/8), 61–83.
- Anscombe, G. E. M. (1978). On the source of the authority of the state. *Ratio* 20(1), 1–28.
- Appiah, K. A. and D. Kodzi (2017). Interview with Kwame Anthony Appiah – Pt. I. *Oxford Review of Books*.
- Aristotle (1984). *Metaphysics*. In J. Barnes (Ed.), *The Complete Works of Aristotle: the Revised Oxford Translation*. Princeton University Press.
- Armour-Garb, B. (2004). Diagnosing dialetheism. In G. Priest, J. Beall, and B. P. Armour-Garb (Eds.), *The Law of Non-Contradiction: New Philosophical Essays*, pp. 113–25. Oxford University Press.
- Armour-Garb, B. and J. A. Woodbridge (2015). *Pretense and Pathology: Philosophical Fictionalism and its Applications*. Cambridge University Press.
- Ayer, A. J. (1936). *Language, Truth, and Logic*. London: V. Gollancz.
- Azzouni, J. (2018). Pathological pretending. *Analysis* 78(4), 692–703.
- Baier, A. (1967). Nonsense. In P. Edwards (Ed.), *The Encyclopedia of Philosophy, volumes 5 and 6*. Collier-MacMillan.
- Balcerak Jackson, B. (2007). Beyond logical form. *Philosophical Studies* 132(2), 347–380.
- Balcerak Jackson, B. (2009). Understanding and semantic structure: Reply to Timothy Williamson. *Proceedings of the Aristotelian Society* 109(1pt3), 337–343.
- Beall, J. (2018). The simple argument for subclassical logic. *Philosophical Issues* 28(1), 30–54.
- Beall, J. and J. Murzi (2013). Two flavors of Curry’s paradox. *Journal of Philosophy* 110(3), 143–165.

- Beall, J. and G. Restall (2005). *Logical Pluralism*. Oxford University Press.
- Belnap, N. (1962). Tonk, plonk and plink. *Analysis* 22(6), 130–134.
- Bergamaschi Ganapini, M. (2019). Belief’s minimal rationality. *Philosophical Studies*, 1–20.
- Berker, S. (2015). Coherentism via graphs. *Philosophical Issues* 25(1), 322–352.
- Besson, C. (2009). Understanding the logical constants and dispositions. *The Baltic International Yearbook of Cognition, Logic and Communication* 5, 1–24.
- Besson, C. (2012). Logical knowledge and ordinary reasoning. *Philosophical Studies* 158(1), 59–82.
- Besson, C. (2019a). Knowledge of logical generality and the possibility of deductive reasoning. In T. Chan and A. Nes (Eds.), *Inference and consciousness. Routledge Studies in Contemporary Philosophy*, pp. 172–196. New York, USA: Routledge.
- Besson, C. (2019b). Logical expressivism and Carroll’s regress. *Royal Institute of Philosophy Supplement* 86, 35–62.
- Bilgrami, A. (1998). Self-knowledge and resentment. In C. Wright, B. C. Smith, and C. Macdonald (Eds.), *Knowing Our Own Minds*. Clarendon Press.
- Bilgrami, A. (2012). *Self-Knowledge and Resentment*. Harvard University Press.
- Blake-Turner, C. and G. Russell (2018). Logical pluralism without the normativity. *Synthese*, 1–19.
- Bochvar, D. A. and M. Bergmann (1937/1981). On a three-valued logical calculus and its application to the analysis of the paradoxes of the classical extended functional calculus. *History and Philosophy of Logic* 2(1-2), 87–112.
- Boghossian, P. (2001). How are objective epistemic reasons possible? *Philosophical Studies* 106(1), 1–40.
- Boghossian, P. (2003). Blind reasoning. *Aristotelian Society Supplementary Volume* 77(1), 225–248.
- Boghossian, P. (2011). Williamson on the a priori and the analytic. *Philosophy and Phenomenological Research* 82(2), 488–497.
- Boghossian, P. (2012). Inferentialism and the epistemology of logic: Reflections on Casalegno and Williamson. *Dialectica* 66(2), 221–236.
- Boghossian, P. (2014). What is inference? *Philosophical Studies* 169(1), 1–18.
- BonJour, L. Analytic philosophy and the nature of thought (ms.).
- BonJour, L. (1998). *In Defense of Pure Reason*. Cambridge University Press.
- Boole, G. (1854). *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*. Dover Constable.

- Boyle, M. (2009). Two kinds of self-knowledge. *Philosophy and Phenomenological Research* 78(1), 133–164.
- Boyle, M. (2016). Additive theories of rationality: A critique. *European Journal of Philosophy* 24(3), 527–555.
- Boyle, M. (2020). Kant on logic and the laws of the understanding. In S. Miguens (Ed.), *The Logical Alien at 20*. Harvard University Press.
- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press.
- Brandom, R. (2018). From logical expressivism to expressivist logic: Sketch of a program and some implementations. *Philosophical Issues* 28(1), 70–88.
- Briggs, R. A. (2019). Conditionals. In R. Pettigrew and J. Weisberg (Eds.), *The Open Handbook of Formal Epistemology*, pp. 543–590. PhilPapers Foundation.
- Brîncuș, C. C. and I. D. Toader (2018). Categoricity and negation. a note on Kripke’s affirmativism. In *The Logica Yearbook 2018*, pp. 57–66. London: College Publications.
- Broome, J. (1999). Normative requirements. *Ratio* 12(4), 398–419.
- Broome, J. (2013). *Rationality Through Reasoning*. Wiley-Blackwell.
- Broome, J. (2014). Comments on Boghossian. *Philosophical Studies* 169(1), 19–25.
- Brouwer, L. E. J. (1951/1981). *Brouwer’s Cambridge Lectures on Intuitionism*. Cambridge University Press.
- Brouwer, L. E. J., M. van Atten, and G. Sundholm (1908/2017). L.E.J. Brouwer’s ‘Unreliability of the Logical Principles’: A new translation, with an introduction. *History and Philosophy of Logic* 38(1), 24–47.
- Burge, T. (1977). Belief *de re*. *Journal of Philosophy* 74(6), 338–362.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4(1), 73–122.
- Burge, T. (1986). Intellectual norms and foundations of mind. *Journal of Philosophy* 83(December), 697–720.
- Burge, T. (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society* 96(1), 91–116.
- Burge, T. (1998a). Frege on knowing the foundation. *Mind* 107(426), 305–347.
- Burge, T. (1998b). Reason and the first person. In C. Wright, B. C. Smith, and C. Macdonald (Eds.), *Knowing Our Own Minds*. Oxford University Press.

- Burge, T. (2007). Postscript to 'Belief *de re*'. In *Foundations of Mind*, pp. 65–81. Oxford University Press.
- Button, T. and S. Walsh (2018). *Philosophy and Model Theory*. Oxford, UK: Oxford University Press.
- Camp, E. (2004). The generality constraint and categorial restrictions. *Philosophical Quarterly* 54(215), 209–231.
- Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research* 78(2), 275–311.
- Cappelen, H. (2013). Nonsense and illusions of thought. *Philosophical Perspectives* 27(1), 22–50.
- Carnap, R. (1943). *Formalization of Logic*. Cambridge: Mass., Harvard University Press.
- Carroll, L. (1895). What the Tortoise said to Achilles. *Mind* 4(14), 278–280.
- Chalmers, D. J. (2011a). Revisability and conceptual change in "Two Dogmas of Empiricism". *Journal of Philosophy* 108(8), 387–415.
- Chalmers, D. J. (2011b). Verbal disputes. *Philosophical Review* 120(4), 515–566.
- Choi, S. and M. Fara (2018). Dispositions. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.
- Christensen, D. (2004). *Putting Logic in its Place: Formal Constraints on Rational Belief*. Oxford University Press.
- Church, A. (1944). Formalization of logic. *Philosophical Review* 53(5), 493.
- Church, A. (1956). *Introduction to Mathematical Logic*. Princeton: Princeton University Press.
- Cohnitz, D. and C. Nicolai. How to adopt a logic (ms.).
- Collins, J. (2007). Syntax, more or less. *Mind* 116(464), 805–850.
- Collins, J. (2015). New thinking about propositions. *Philosophical Quarterly* 65(261), 842–845.
- Conant, J. (1992). The search for logically alien thought: Descartes, Kant, Frege, and the Tractatus. *Philosophical Topics* 20(1), 115–180.
- Conant, J. (2002). The method of the Tractatus. In E. H. Reck (Ed.), *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*. Oxford University Press.
- Cotnoir, A. J. (2018). Logical nihilism. In J. Wyatt, N. J. Pedersen, and N. Kellen (Eds.), *Pluralisms in Truth and Logic*. Palgrave MacMillan.
- Crimmins, M. (1998). Hesperus and Phosphorus: Sense, pretense, and reference.

- Philosophical Review* 107(1), 1–47.
- Dasgupta, S. (2018). Realism and the absence of value. *Philosophical Review* 127(3), 279–322.
- Davidson, D. (1967). Truth and meaning. *Synthese* 17(1), 304–323.
- Davidson, D. (1973). Radical interpretation. *Dialectica* 27(3-4), 313–328.
- Davidson, D. (1985). Incoherence and irrationality. *Dialectica* 39(4), 345–354.
- Davidson, D. (1999). The emergence of thought. *Erkenntnis* 51(1), 7–17.
- Davidson, D. (2004). *Problems of Rationality*. Oxford University Press.
- Davies, M. (1981). *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. Routledge & Kegan Paul.
- Davies, M. (1989). Tacit knowledge and subdoxastic states. In A. George (Ed.), *Reflections on Chomsky*. Blackwell.
- Descartes, R. (1996). *Meditations on First Philosophy: With Selections From the Objections and Replies*. Cambridge University Press.
- Diamond, C. (1981). What nonsense might be. *Philosophy* 56(215), 5–22.
- Diamond, C. (1988). Throwing away the ladder. *Philosophy* 63(243), 5–27.
- Diamond, C. (2000). Ethics, imagination and the method of Wittgenstein's Tractatus. In A. Crary and R. J. Read (Eds.), *The New Wittgenstein*, pp. 149–173. Routledge.
- Dicher, B. (2016). A proof-theoretic defence of meaning-invariant logical pluralism. *Mind* 125(499), 727–757.
- Dickie, I. (2010). Negation, anti-realism, and the denial defence. *Philosophical Studies* 150(2), 161–185.
- Dickie, I. (2015). *Fixing Reference*. Oxford University Press.
- Dickie, I. (2017). Everybody needs to know? *Philosophical Studies* 174(10), 2571–2583.
- Dogramaci, S. (2013). Intuitions for inferences. *Philosophical Studies* 165(2), 371–399.
- Dogramaci, S. (2015). Communist conventions for deductive reasoning. *Noûs* 49(4), 776–799.
- Došen, K. (1989). Logical constants as punctuation marks. *Notre Dame Journal of Formal Logic* 30(3), 362–381.
- Dummett, M. A. E. (1959). Truth. *Proceedings of the Aristotelian Society* 59(1), 141–62.
- Dummett, M. A. E. (1973). *Frege: Philosophy of Language*. London: Duckworth.
- Dummett, M. A. E. (1974). The justification of deduction. In M. A. E. Dummett (Ed.), *Truth and Other Enigmas*. Oxford University Press.

- Dummett, M. A. E. (1975). What is a theory of meaning? (I). In S. Guttenplan (Ed.), *Mind and Language*. Oxford University Press.
- Dummett, M. A. E. (1991). *The Logical Basis of Metaphysics*. Harvard University Press.
- Dummett, M. A. E. (1994). *Origins of Analytical Philosophy*. Harvard University Press.
- Dunn, J. M. (1976). Intuitive semantics for first-degree entailments and ‘coupled trees’. *Philosophical Studies* 29(3), 149–168.
- Edgington, D. (1995). On conditionals. *Mind* 104(414), 235–329.
- Enderton, H. B. (1972). *A Mathematical Introduction to Logic*. New York: Academic Press.
- Engel, P. (2005). Logical reasons. *Philosophical Explorations* 8(1), 21–38.
- Enoch, D. (2006). Agency, shmagency: Why normativity won’t come from what is constitutive of action. *Philosophical Review* 115(2), 169–198.
- Etchemendy, J. (1990). *The Concept of Logical Consequence*. Harvard University Press.
- Evans, G. (1976). Semantic structure and logical form. In G. Evans and J. McDowell (Eds.), *Truth and Meaning*, pp. 49–75. Oxford University Press.
- Evans, G. (1978). Can there be vague objects? *Analysis* 38(4), 208.
- Evans, G. (1981). Semantic theory and tacit knowledge. In S. Holtzman and C. M. Leich (Eds.), *Wittgenstein: To Follow a Rule*. Routledge.
- Evans, G. (1982). *The Varieties of Reference*. Oxford University Press.
- Evnine, S. J. (2001). The universality of logic: On the connection between rationality and logical ability. *Mind* 110(438), 335–367.
- Feferman, S. (1984). Toward useful type-free theories (I). *Journal of Symbolic Logic* 49(1), 75–111.
- Field, H. (1973). Theory change and the indeterminacy of reference. *Journal of Philosophy* 70(14), 462–481.
- Field, H. (1978). Mental representation. *Erkenntnis* 13(July), 9–61.
- Field, H. (1998). Epistemological nonfactualism and the a prioricity of logic. *Philosophical Studies* 92(1), 1–24.
- Field, H. (2009). What is the normative role of logic? *Aristotelian Society Supplementary Volume* 83(1), 251–268.
- Field, H. (2017). Egocentric content. *Noûs* 51(3), 521–546.
- Fine, K. (1975). Vagueness, truth and logic. *Synthese* 30(3-4), 265–300.

- Fine, K. (2014). Truth-maker semantics for intuitionistic logic. *Journal of Philosophical Logic* 43(2-3), 549–577.
- Finn, S. (2019a). The adoption problem and anti-exceptionalism about logic. *Australasian Journal of Logic* 16(7), 231.
- Finn, S. (2019b). Limiting logical pluralism. *Synthese*.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Foot, P. (2001). *Natural Goodness*. Oxford University Press.
- Frege, G. (1879/1967). Begriffsschrift, eine der arithmetischen nachgebildete formelsprache des reinen denkens. In J. van Heijenoort (Ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge: Harvard University Press.
- Frege, G. (1892/1960b). On sense and reference. In P. Geach and M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege*. Basil Blackwell.
- Frege, G. (1897/1979). Logic. In P. Long and R. M. White (Eds.), *Posthumous Writings*. Blackwell.
- Frege, G. (1919/1960a). Negation. In P. Geach and M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege*. Basil Blackwell.
- Frege, G. and J. L. Austin (1884/1953). *Foundations of Arithmetic*. Blackwell.
- Frege, G., P. A. Ebert, and M. Rossberg (1893/2013). *Basic Laws of Arithmetic*. Oxford University Press.
- Fricker, E. (1982). Semantic structure and speakers' understanding. *Proceedings of the Aristotelian Society* 83(1), 49–66.
- Geach, P. T. (1976). Why logic matters. In H. Lewis (Ed.), *Contemporary British Philosophy, Fourth Series*. George Allen and Unwin.
- Gentzen, G. (1969). *The Collected Papers of Gerhard Gentzen*. Amsterdam: North-Holland Pub. Co.
- Glock, H. (2015). Nonsense made intelligible. *Erkenntnis* 80(1), 111–136.
- Glüer, K. and A. Wikforss (2009). Against content normativity. *Mind* 118(469), 31–70.
- Goddard, L. (1970). Nonsignificance. *Australasian Journal of Philosophy* 48(1), 10–16.
- Goddard, L. and R. Routley (1973). *The Logic of Significance and Context*. New York: Wiley.
- Goldstick, D. (1974). Against 'categories'. *Philosophical Studies* 26(5-6), 337–356.

- Goodship, L. (1996). On dialethism. *Australasian Journal of Philosophy* 74(1), 153–161.
- Grice, H. P. (1957). Meaning. *Philosophical Review* 66(3), 377–388.
- Grim, P. (2004). What is a contradiction? In G. Priest, J. Beall, and B. P. Armour-Garb (Eds.), *The Law of Non-Contradiction : New Philosophical Essays*, pp. 49–72. Oxford University Press.
- Haack, S. (1996). *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. University of Chicago Press.
- Habib, A. (2018). Promises. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 ed.). Metaphysics Research Lab, Stanford University.
- Hacker-Wright, J. (2012). Ethical naturalism and the constitution of agency. *Journal of Value Inquiry* 46(1), 13–23.
- Hale, B. (1999). On some arguments for the necessity of necessity. *Mind* 108(429), 23–52.
- Hale, B. (2002). Basic logical knowledge. *Royal Institute of Philosophy Supplement* 51, 279–304.
- Halldén, S. (1949). *The Logic of Nonsense*. Uppsala Universitets årsskrift.
- Harman, G. (1984). Logic and reasoning. *Synthese* 60(1), 107–127.
- Harman, G. (1986). *Change in View*. MIT Press.
- Hart, H. L. A. (1961). *The Concept of Law*. Oxford University Press.
- Haslanger, S. (2000). Gender and race: (what) are they? (what) do we want them to be? *Noûs* 34(1), 31–55.
- Hattiangadi, A. (2006). Is meaning normative? *Mind and Language* 21(2), 220–240.
- Hazen, A. P. and F. J. Pelletier (2020). K3, Ł3, LP, RM3, A3, FDE, M: How to make many-valued logics work for you. In H. Omori and H. Wansing (Eds.), *New Essays on Belnap-Dunn Logic. Synthese Library (Studies in Epistemology Logic, Methodology, and Philosophy of Science)*, vol 418. Springer, Cham.
- Heal, J. (1989). *Fact and Meaning: Quine and Wittgenstein on Philosophy of Language*. Blackwell.
- Heal, J. (1994). Semantic holism: Still a good buy. *Proceedings of the Aristotelian Society* 94(325).
- Heal, J. (1999). Thoughts and holism: Reply to Cohen. *Analysis* 59(2), 71–78.
- Heal, J. (2017). Radical interpretation. In *Blackwell Companion to Philosophy of Language*, pp. 299–317. Wiley Blackwell.
- Hegel, G. W. F. and G. di Giovanni (1816/2010). *The Science of Logic*. Cambridge

- University Press.
- Hegel, G. W. F., T. F. Geraets, W. A. Suchting, and H. S. Harris (1817/1991). *The Encyclopedia Logic: Part I of the Encyclopedia of Philosophical Sciences with the Zusätze*. Hackett Publishing Co.
- Higginbotham, J. (1989). Knowledge of reference. In A. George (Ed.), *Reflections on Chomsky*, pp. 153–74. Blackwell.
- Higginbotham, J. and G. Segal (1994). Priorities in the philosophy of thought. *Aristotelian Society Supplementary Volume 68*(1), 85–130.
- Hjortland, O. T. (2013). Logical pluralism, meaning-variance, and verbaldisputes. *Australasian Journal of Philosophy 91*(2), 355–373.
- Hjortland, O. T. (2017). Anti-exceptionalism about logic. *Philosophical Studies 174*(3), 631–658.
- Hlobil, U. (2015). There are diachronic norms of rationality. *Thought: A Journal of Philosophy 4*(1), 38–45.
- Hlobil, U. (2016). A nonmonotonic sequent calculus for inferentialist expressivists. In P. Arazim and M. Danák (Eds.), *The Logica Yearbook 2015*, pp. 87–105. College Publications.
- Horvath, J. (2020). Understanding as a source of justification. *Mind 129*(514), 509–534.
- Humberstone, L. (2000). The revival of rejective negation. *Journal of Philosophical Logic 29*(4), 331–381.
- Humberstone, L. (2011). *The Connectives*. MIT Press.
- Husserl, E. and D. Cairns (1929/1969). *Formal and Transcendental Logic*. The Hague: Martinus Nijhoff.
- Husserl, E. and J. N. Findlay (1899/1970). *Logical Investigations*. Routledge & Kegan Paul.
- Ichikawa, J. J. and B. W. Jarvis (2013). *The Rules of Thought*. Oxford University Press.
- Incurvati, L. and J. J. Schlöder (2017). Weak rejection. *Australasian Journal of Philosophy 95*(4), 741–760.
- Incurvati, L. and P. Smith (2010). Rejection and valuations. *Analysis 70*(1), 3–10.
- Jenkins, K. (2016). Amelioration and inclusion: Gender identity and the concept of woman. *Ethics 126*(2), 394–421.
- Jeshion, R. (2001). Frege’s notions of self-evidence. *Mind 110*(440), 937–976.
- Kant, I., P. Guyer, and E. Mathews (1790/2000). *Critique of the Power of Judgment*.

- Cambridge University Press.
- Kant, I., P. Guyer, and A. Wood (1781/1998). *Critique of Pure Reason*. Cambridge: Cambridge University Press.
- Kant, I. and J. M. Young (1800/1992). Jäsche Logic. In *Lectures on Logic*. Cambridge University Press.
- Keefe, R. and P. Smith (1996). *Vagueness: A Reader*. MIT Press.
- Kimhi, I. (2018). *Thinking and Being*. Harvard University Press.
- Kissel, T. K. and S. Shapiro (2017). Logical pluralism and normativity. *Inquiry: An Interdisciplinary Journal of Philosophy*, 1–22.
- Kleene, S. C. (1952). *Introduction to Metamathematics*. North Holland.
- Kolodny, N. and J. MacFarlane (2010). Ifs and oughts. *Journal of Philosophy* 107(3), 115–143.
- Koreň, L. (2018). Propositional contents and the logical space. In O. Beran, V. Kolman, and L. Koreň (Eds.), *From Rules to Meanings: New Essays on Inferentialism*. Routledge.
- Kornblith, H. (1993). Epistemic normativity. *Synthese* 94(3), 357–376.
- Korsgaard, C. M. (1983). Two distinctions in goodness. *Philosophical Review* 92(2), 169–195.
- Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge University Press.
- Korsgaard, C. M. (2008). *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford University Press.
- Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.
- Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy* 72(19), 690–716.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press.
- Kripke, S. (2015). Yet another dogma of empiricism. *Philosophy and Phenomenological Research* 91(2), 381–385.
- Kroon, F. (2018). Pushing the boundaries of pretence. *Analysis* 78(4), 703–712.
- Kürbis, N. (2016). Some comments on Ian Rumfitt’s bilateralism. *Journal of Philosophical Logic* 45(6), 623–644.
- Kürbis, N. (2019). *Proof and Falsity: A Logical Investigation*. Cambridge, UK: Cambridge University Press.
- Kwong, J. M. C. (2007). Is conceptual atomism a plausible theory of concepts?

- Southern Journal of Philosophy* 45(3), 413–434.
- Lambek, J. (1958). The mathematics of sentence structure. *The American Mathematical Monthly* 65(3).
- Lavin, D. (2004). Practical reason and the possibility of error. *Ethics* 114(3), 424–457.
- Leech, J. (2015). Logic and the laws of thought. *Philosophers' Imprint* 15.
- Leech, J. (2017). The normativity of Kant's logical laws. *History of Philosophy Quarterly* 34(4).
- Leibniz, G. W. (1765/1990). *New Essays Concerning Human Understanding*. The Continental Rationalists. Intelelex Corporation.
- Lepore, E. and K. Ludwig (2002). What is logical form? In G. Preyer and G. Peter (Eds.), *Logical Form and Language*, pp. 54–90. Clarendon Press.
- Lewis, D. K. (1971). Immodest inductive methods. *Philosophy of Science* 38(1), 54–63.
- Lewis, D. K. (1982). Logic for equivocators. *Noûs* 16(3), 431–441.
- Lewis, D. K. (1986). *On the Plurality of Worlds*. Wiley-Blackwell.
- Lewis, D. K. (1997). Finkish dispositions. *Philosophical Quarterly* 47(187), 143–158.
- Lindeman, K. (2017). Constitutivism without normative thresholds. *Journal of Ethics and Social Philosophy* 3(XII), 231–258.
- Locke, J. (1689). *An Essay Concerning Human Understanding*. Oxford University Press.
- Lott, M. (2014). Why be a good human being? Natural goodness, reason, and the authority of human nature. *Philosophia* 42(3), 761–777.
- Ludlow, P. (2003). Externalism, logical form, and linguistic intentions. In A. Barber (Ed.), *Epistemology of Language*, pp. 399–414. Oxford: Oxford University Press.
- Łukasiewicz, J. (1930/1967). Philosophical remarks on many-valued systems of propositional logic. In S. McCall (Ed.), *Polish Logic, 1920-1939*. Oxford: Clarendon Press.
- Łukasiewicz, J. (1937/1970). In defence of logic. In L. Borkowski (Ed.), *Jan Łukasiewicz. Selected Works*. North-Holland Publishing Company.
- Lu-Adler, H. (2017). Kant and the normativity of logic. *European Journal of Philosophy* 25(2), 207–230.
- MacFarlane, J. (2000). *What Does It Mean to Say That Logic is Formal?* Ph. D. thesis, University of Pittsburgh.
- MacFarlane, J. (2002). Frege, Kant, and the logic in logicism. *Philosophical Re-*

- view* 111(1), 25–65.
- MacFarlane, J. (2004). In what sense (if any) is logic normative for thought? (ms.).
- Maddy, P. (2002). A naturalistic look at logic. *Proceedings and Addresses of the American Philosophical Association* 76(2), 61–90.
- Maddy, P. (2014). A second philosophy of logic. In P. Rush (Ed.), *The Metaphysics of Logic*, pp. 93–108. Cambridge University Press.
- Magidor, O. (2013). *Category Mistakes*. Oxford University Press.
- Marcus, E. (2020). Inference as consciousness of necessity. *Analytic Philosophy*.
- McDowell, J. (1976). Truth-conditions, bivalence, and verification. In G. Evans and J. McDowell (Eds.), *Truth and Meaning*. Clarendon Press.
- McDowell, J. (1994). *Mind and World*. Cambridge: Harvard University Press.
- McDowell, J. (1997a). Another plea for modesty. In R. G. Heck (Ed.), *Language, Thought, and Logic: Essays in Honour of Michael Dummett*. Oxford University Press.
- McDowell, J. (1997b). Brandom on representation and inference. *Philosophy and Phenomenological Research* 57(1), 157–162.
- McDowell, J. (2005). Motivating inferentialism: Comments on Making It Explicit, ch. 2. *Pragmatics and Cognition* 13(1), 121–140.
- McFetridge, I. (1990). Logical necessity: Some issues. In J. Haldane and R. Scruton (Eds.), *Logical Necessity and other essays*. The Aristotelian Society.
- McGee, V. (1985). A counterexample to modus ponens. *Journal of Philosophy* 82(9), 462–471.
- Merritt, M. M. (2015). Varieties of reflection in Kant’s logic. *British Journal for the History of Philosophy* 23(3), 478–501.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press.
- Milne, P. (2009). What is the normative role of logic? *Aristotelian Society Supplementary Volume* 83(1), 269–298.
- Moore, A. W. (2000). *Points of View*. Oxford University Press.
- Moore, A. W. (2003a). Ineffability and nonsense. *Aristotelian Society Supplementary Volume* 77(1), 169–193.
- Moore, A. W. (2003b). *Noble in Reason, Infinite in Faculty: Themes and Variations in Kants Moral and Religious Philosophy*. Routledge.
- Moore, A. W. (2018). *The Infinite* (Third Edition ed.). Routledge.
- Moosavi, P. (2019). From biological functions to natural goodness. *Philosophers’*

Imprint 19.

- Moran, R. A. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.
- Murzi, J. (2020). Classical harmony and separability. *Erkenntnis* 85(2), 391–415.
- Murzi, J. and O. T. Hjortland (2009). Inferentialism and the categoricity problem: Reply to Raatikainen. *Analysis* 69(3), 480–488.
- Murzi, J. and F. Steinberger (2013). Is logical knowledge dispositional? *Philosophical Studies* 166(S1), 165–183.
- Murzi, J. and F. Steinberger (2017). Inferentialism. In *Blackwell Companion to Philosophy of Language*, pp. 197–224. Wiley Blackwell.
- Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.
- Nunez, T. (2018). Logical mistakes, logical aliens, and the laws of Kant’s pure general logic. *Mind* 128(512), 1149–1180.
- Oliver, A. (2010). The matter of form : Logic’s beginnings. In T. J. Smiley, J. Lear, and A. Oliver (Eds.), *The Force of Argument: Essays in Honor of Timothy Smiley*, pp. 165–185. London: Routledge.
- Padró, R. (2015). *What the Tortoise Said to Kripke: the Adoption Problem and the Epistemology of Logic*. Ph. D. thesis, CUNY Academic Works.
- Paoli, F. (2007). Implicational paradoxes and the meaning of logical constants. *Australasian Journal of Philosophy* 85(4), 553–579.
- Pap, A. (1960). Types and meaninglessness. *Mind* 69(273), 41–54.
- Parfit, D. (1997). Reasons and motivation. *Aristotelian Society Supplementary Volume* 71(1), 99–130.
- Peacocke, C. (1986). Explanation in computational psychology: Language, perception and level 1.5. *Mind and Language* 1(2), 101–123.
- Peacocke, C. (1987). Understanding logical constants: A realist’s account. In *Proceedings of the British Academy, Volume 73: 1987*, pp. 153.
- Peacocke, C. (1992). *A Study of Concepts*. MIT Press.
- Peacocke, C. (1993). Proof and truth. In J. Haldane and C. Wright (Eds.), *Reality, Representation, and Projection*, pp. 165–190. Oxford University Press.
- Peacocke, C. (2004). *Realm of Reason*. Oxford University Press UK.
- Peacocke, C. (2008). *Truly Understood*. Oxford University Press.
- Pietarinen, A. (2005). Cultivating habits of reason: Peirce and the logica utens versus logica docens distinction. *History of Philosophy Quarterly* 22(4), 357–372.
- Price, H. (1990). Why ‘not’? *Mind* 99(394), 221–238.

- Priest, G. (1979). Two dogmas of Quineanism. *Philosophical Quarterly* 29(117), 289–301.
- Priest, G. (2005). *Doubt Truth to Be a Liar*. Oxford University Press.
- Priest, G. (2006). *In Contradiction: A Study of the Transconsistent*. Oxford University Press.
- Priest, G. (2016). Logical disputes and the a priori. *Logique et Analyse* 59(236).
- Prior, A. N. (1960). The runabout inference-ticket. *Analysis* 21(2), 38.
- Putnam, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science* 7, 131–193.
- Putnam, H. (1978). There is at least one a priori truth. *Erkenntnis* 13(1), 153–170.
- Putnam, H. (1994). Rethinking mathematical necessity. In J. Conant (Ed.), *Words and Life*, pp. 245–263.
- Quine, W. V. O. (1948). On what there is. *Review of Metaphysics* 2(5), 21–38.
- Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review* 60(1), 20–43.
- Quine, W. V. O. (1960). *Word and Object*. MIT Press.
- Quine, W. V. O. (1961). Notes on the theory of reference. In *From a Logical Point of View*. Harper Torchbooks.
- Quine, W. V. O. (1970a). Methodological reflections on current linguistic theory. *Synthese* 21(3-4), 386–398.
- Quine, W. V. O. (1970b). *Philosophy of Logic*. Harvard University Press.
- Quine, W. V. O. (1974). *The Roots of Reference*. LaSalle, Ill., Open Court.
- Quine, W. V. O. (1992). *Pursuit of Truth, Revised Edition*. Cambridge: Mass.: Harvard University Press.
- Raatikainen, P. (2008). On rules of inference and the meanings of logical constants. *Analysis* 68(4), 282–287.
- Rabinowicz, W. and T. Rønnow-Rasmussen (2000). A distinction in value: Intrinsic and for its own sake. *Proceedings of the Aristotelian Society* 100(1), 33–51.
- Raffman, D. (2014). *Unruly Words: A Study of Vague Language*. Oxford University Press.
- Railton, P. (1997). On the hypothetical and non-hypothetical in reasoning about belief and action. In G. Cullity and B. N. Gaut (Eds.), *Ethics and Practical Reason*, pp. 53–79. Oxford University Press.
- Rattan, G. (2002). Tacit knowledge of grammar: A reply to Knowles. *Philosophical Psychology* 15(2), 135–154.

- Rattan, G. (2014). Review of *The Rules of Thought*. *Notre Dame Philosophical Reviews*.
- Rattan, G. and A. Wikforss (2017). Is understanding epistemic in nature? *Pacific Philosophical Quarterly* 98(2).
- Reimer, M. (2012). Davidsonian holism in recent philosophy of psychiatry. In G. Preyer (Ed.), *Donald Davidson on Truth, Meaning, and the Mental*. Oxford University Press.
- Restall, G. (1999). Negation in relevant logics (how I stopped worrying and learned to love the Routley Star). In D. M. Gabbay and H. Wansing (Eds.), *What is Negation?*, pp. 53–76. Kluwer Academic Publishers.
- Restall, G. (2000). *An Introduction to Substructural Logics*. Routledge.
- Restall, G. (2004). Laws of non-contradiction, laws of the excluded middle, and logics. In G. Priest, J. C. Beall, and B. Armour-Garb (Eds.), *The Law of Non-Contradiction: New Philosophical Essays*. Clarendon Press.
- Restall, G. (2005). Multiple conclusions. In P. Hájek, L. Valdés-Villanueva, and D. Westerståhl (Eds.), *Logic, Methodology and Philosophy of Science*. College Publications.
- Restall, G. (2013). Assertion, denial and non-classical theories. In F. Berto, E. Mares, K. Tanaka, and F. Paoli (Eds.), *Paraconsistency: Logic and Applications*, pp. 81–99. Springer.
- Restall, G. (2014). Pluralism and proofs. *Erkenntnis* 79(S2), 279–291.
- Richard, M. (2000). Semantic pretense. In T. Hofweber and A. Everett (Eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence*, pp. 205–32. CSLI Publications.
- Richard, M. (2013). *Context and the Attitudes: Meaning in Context, Volume 1*. Oxford University Press.
- Ripley, D. (2015). Embedding denial. In C. Caret and O. Hjortland (Eds.), *Foundations of Logical Consequence*, pp. 289–309. Oxford University Press.
- Rödl, S. (2006). Logical form as a relation to the object. *Philosophical Topics* 34(1/2), 345–369.
- Rödl, S. (2018). *Self-Consciousness and Objectivity: An Introduction to Absolute Idealism*. Harvard University Press.
- Ross, J. (1992). Immaterial aspects of thought. *Journal of Philosophy* 89(3), 136–150.
- Routley, R. (1966). On a significance theory. *Australasian Journal of Philosophy* 44(2), 172–209.
- Routley, R. (1969). The need for nonsense. *Australasian Journal of Philosophy* 47(3),

- 367–384.
- Rumfitt, I. (1997). The categoricity problem and truth-value gaps. *Analysis* 57(4), 223–236.
- Rumfitt, I. (2000). Yes and no. *Mind* 109(436), 781–823.
- Rumfitt, I. (2002). Unilateralism disarmed: A reply to Dummett and Gibbard. *Mind* 111(442), 305–322.
- Rumfitt, I. (2011). Inference, deduction, logic. In J. Bengson and M. A. Moffett (Eds.), *Knowing How: Essays on Knowledge, Mind, and Action*, pp. 334. Oxford University Press.
- Rumfitt, I. (2012). On a neglected path to intuitionism. *Topoi* 31(1), 101–109.
- Rumfitt, I. (2014). Truth and meaning. *Aristotelian Society Supplementary Volume* 88(1), 21–55.
- Rumfitt, I. (2015). *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford University Press.
- Rumfitt, I. (2017). Against harmony. In B. Hale, C. Wright, and A. Miller (Eds.), *The Blackwell Companion to the Philosophy of Language*. Blackwell.
- Rumfitt, I. (2018). Reply to Crispin Wright and Richard Zach. *Philosophical Studies* 175(8), 2091–2103.
- Russell, B. (1912). *The Problems of Philosophy*. Home University Library.
- Russell, B. (1919). The Philosophy of Logical Atomism, lectures 5-6. *The Monist* 29(2), 190–222.
- Russell, B. (1923). Vagueness. *Australasian Journal of Philosophy* 1(2), 84–92.
- Russell, G. (2008). One true logic? *Journal of Philosophical Logic* 37(6), 593–611.
- Russell, G. (2018a). Logical nihilism: Could there be no logic? *Philosophical Issues* 28(1), 308–324.
- Russell, G. (2018b). Varieties of logical consequence by their resistance to logical nihilism. In J. Wyatt, N. J. Pedersen, and N. Kellen (Eds.), *Pluralisms in Truth and Logic*. Palgrave MacMillan.
- Russell, G. (2019). Deviance and vice: Strength as a theoretical virtue in the epistemology of logic. *Philosophy and Phenomenological Research* 99(3), 548–563.
- Russell, G. (2020). Logic isn't normative. *Inquiry: An Interdisciplinary Journal of Philosophy* 63(3-4), 371–388.
- Ryle, G. (1949). *The Concept of Mind*. Hutchinson & Co.
- Rysiew, P. (2002). Reid and epistemic naturalism. *Philosophical Quarterly* 52(209), 437–456.

- Sainsbury, R. M. (2002). What logic should we think with? *Royal Institute of Philosophy Supplement* 51, 1–17.
- Sambin, G., G. Battilotti, and C. Faggian (2000). Basic logic: Reflection, symmetry, visibility. *Journal of Symbolic Logic* 65(3), 979–1013.
- Schechter, J. and D. Enoch (2006). Meaning and justification: The case of modus ponens. *Noûs* 40(4), 687–715.
- Schroeter, L. and F. Schroeter (2009). A third way in metaethics. *Noûs* 43(1), 1–30.
- Sellars, W. (1974). Meaning as functional classification. *Synthese* 27(3-4), 417–437.
- Shapiro, L. (2011). Deflating logical consequence. *Philosophical Quarterly* 61(243), 320–342.
- Shapiro, L. (2015). Naive structure, contraction and paradox. *Topoi* 34(1), 75–87.
- Shapiro, S. (2014). *Varieties of Logic*. Oxford University Press.
- Shaw, J. R. (2014). What is a truth-value gap? *Linguistics and Philosophy* 37(6), 503–534.
- Shoesmith, D. J. and T. J. Smiley (1978). *Multiple-Conclusion Logic*. Cambridge University Press.
- Sibajiban (1964). The self-contradictory and the inconceivable. *Analysis* 24(Suppl-2), 99–103.
- Silverstein, M. (2015). The shmagency question. *Philosophical Studies* 172(5), 1127–1142.
- Smiley, T. (1996). Rejection. *Analysis* 56(1), 1–9.
- Sorensen, R. A. (2001). *Vagueness and Contradiction*. Oxford University Press.
- Sorensen, R. A. (2002). Meaningless beliefs and Mates’s problem. *American Philosophical Quarterly* 39(2), 169–182.
- Soysal, Z. (2018). Formal analyticity. *Philosophical Studies* 175(11), 2791–2811.
- Sperber, D., F. Cara, and V. Girotto (1995). Relevance theory explains the selection task. *Cognition* 57(1), 31–95.
- Stang, N. (2014). Kant, Bolzano, and the formality of logic. In S. Lapointe and C. Tolley (Eds.), *The New Anti-Kant*, pp. 193–234.
- Steinberger, F. (2011a). What harmony could and could not be. *Australasian Journal of Philosophy* 89(4), 617–639.
- Steinberger, F. (2011b). Why conclusions should remain single. *Journal of Philosophical Logic* 40(3), 333–355.
- Steinberger, F. (2017). Frege and Carnap on the normativity of logic. *Synthese* 194(1), 143–162.

- Steinberger, F. (2019a). Consequence and normative guidance. *Philosophy and Phenomenological Research* 98(2), 306–328.
- Steinberger, F. (2019b). Logical pluralism and logical normativity. *Philosophers' Imprint* 19.
- Steinberger, F. (2019c). Three ways in which logic might be normative. *Journal of Philosophy* 116(1), 5–31.
- Stich, S. (1990). *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, Massachusetts: MIT Press.
- Strawson, P. F. (1950). On referring. *Mind* 59(235), 320–344.
- Stroud, B. (1979). Inference, belief, and understanding. *Mind* 88(350), 179–196.
- Sundholm, G. (2002). A century of inference: 1837–1936. In P. Gärdenfors, J. Woleński, and K. Kijania-Placek (Eds.), *In the Scope of Logic, Methodology and Philosophy of Science: Volume Two of the 11th International Congress of Logic, Methodology and Philosophy of Science, Cracow, August 1999*, pp. 565–580. Dordrecht: Springer Netherlands.
- Suszko, R. (1977). The Fregean axiom and Polish mathematical logic in the 1920s. *Studia Logica* 36(4), 377–380.
- Sylvan, R. (1999). What is that item designated negation? In D. M. Gabbay and H. Wansing (Eds.), *What is Negation?*, pp. 299–324. Kluwer Academic Publishers.
- Tarski, A. (1923/1998). On the primitive term of logistic. In J. Srzednicki and Z. Stachniak (Eds.), *Leśniewski's Systems. Protothetic*. Kluwer Academic Publishers.
- Tarski, A. and J. Corcoran (1966/1986). What are logical notions? *History and Philosophy of Logic* 7(2), 143–154.
- Tenenbaum, S. (2011). Review of Christine Korsgaard's *Self-Constitution*. *Ethics* 121(2), 449–455.
- Tennant, N. (1978). Entailment and proofs. *Proceedings of the Aristotelian Society* 79, 167–189+viii.
- Tennant, N. (1999). Negation, absurdity, and contrariety. In D. Gabbay and H. Wansing (Eds.), *What is Negation?. Applied Logic Series vol 13*. Dordrecht: Springer.
- Tennant, N. (2002). *The Taming of the True*. Oxford University Press.
- Thompson, M. Three degrees of natural goodness (ms.).
- Tolley, C. (2006). Kant on the nature of logical laws. *Philosophical Topics* 34(1/2), 371–407.

- Troelstra, A. (1992). *Lectures on Linear Logic, CSLI Lecture Notes 29*. Stanford: CSLI Publications.
- Valaris, M. (2017). What reasoning might be. *Synthese* 194(6).
- Valaris, M. (2019). Reasoning and deducing. *Mind* 128(511), 861–885.
- van Dalen, D. (2001). Intuitionistic logic. In L. Goble (Ed.), *The Blackwell Guide to Philosophical Logic*. Wiley-Blackwell.
- Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.
- Walton, K. L. (1993). Metaphor and prop oriented make-believe. *European Journal of Philosophy* 1(1), 39–57.
- Warren, J. (2015). Talking with tonkers. *Philosophers' Imprint* 15.
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology* 20(3), 273–281.
- Wedgwood, R. (2006). The normative force of reasoning. *Noûs* 40(4), 660–686.
- Wedgwood, R. (2017). *The Value of Rationality*. Oxford University Press.
- Wiggins, D. (1997). *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford University Press.
- Wikforss, A. (2009). Are there understanding-assent links? *The Baltic International Yearbook of Cognition, Logic and Communication* 5(12).
- Williams, B. (2006). The human prejudice. In *Philosophy as a Humanistic Discipline*. Princeton University Press.
- Williamson, T. (1994). *Vagueness*. Routledge.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.
- Williamson, T. (2006). Conceptual truth. *Aristotelian Society Supplementary Volume* 80(1), 1–41.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Wiley-Blackwell.
- Williamson, T. (2013). *Modal Logic as Metaphysics*. Oxford University Press.
- Williamson, T. (2017). Dummett on the relation between logics and metalogics. In M. Frauchiger (Ed.), *Truth, Meaning, Justification, and Reality: Themes From Dummett*, pp. 153–176. De Gruyter.
- Williamson, T. (2018). Alternative logics and applied mathematics. *Philosophical Issues* 28(1), 399–424.
- Wittgenstein, L. (1921/1961). *Tractatus Logico-Philosophicus (Trans. Pears and McGuinness)*. Routledge.
- Wittgenstein, L. (1953/2009). *Philosophical Investigations, 4th Edition (Trans.*

- Hacker and Schulte*). Wiley-Blackwell.
- Wittgenstein, L. (1969). *On Certainty* (Ed. Anscombe and von Wright). Harper Torchbooks.
- Woleński, J. (2003). Psychologism and metalogic. *Synthese* 137(1-2), 179–193.
- Woods, J. (2018). Pretendism in name only. *Analysis* 78(4), 713–718.
- Wright, C. (1981). Rule-following, objectivity and the theory of meaning. In S. H. Holtzman and C. M. Leich (Eds.), *Wittgenstein: To Follow A Rule*. Routledge.
- Wright, C. (1986). Inventing logical necessity. In J. Butterfield (Ed.), *Language, Mind and Logic*. Cambridge University Press.
- Wright, C. (1987). Theories of meaning and speakers' knowledge. In *Realism, Meaning and Truth*. Blackwell.
- Wright, C. (1993). On an argument on behalf of classical negation. *Mind* 102(405), 123–131.
- Wright, C. (2014). Comment on Paul Boghossian, 'What is Inference'. *Philosophical Studies* 169(1), 27–37.
- Wright, C. (2018). Logical non-cognitivism. *Philosophical Issues* 28(1), 425–450.
- Yablo, S. (2014). *Aboutness*. Princeton University Press.
- Zalabardo, J. L. (2011). Boghossian on inferential knowledge. *Analytic Philosophy* 52(2), 124–139.
- Zwicky, J. (2019). *The Experience of Meaning*. McGill-Queen's University Press.