**Non-cognitivism About Metaphysical Explanation**

**Abstract**

This paper introduces a non-cognitivist account of metaphysical explanation according to which the core function of judgements of the form ⌜x *because* y⌝ is not to state truth-apt beliefs. Instead, their core function is to express attitudes of *commitment* *to*, and *recommendation of the acceptance of* certain norms governing interventional conduct at contexts.

**1. Introduction**

It is common for individuals to make judgements about which things *metaphysically explain* which other things: judgements of the form ⌜x *because* y⌝,[[1]](#footnote-1) where ⌜x⌝ and ⌜y⌝ are sentences, and *‘because’* expresses a particular kind of explanatory connection (i.e. one that is metaphysical rather than causal).[[2]](#footnote-2) In order not to prejudge the issues of this paper, in what follows we use ‘*judgement’* to refer to mental states that are expressed by sincere utterances of ⌜x *because* y⌝, leaving open whether those mental states are cognitive or non-cognitive.

Alongside judgements about which things *causally explain* which other things, judgements of ⌜x *because* y⌝ are a staple of everyday thought and discourse. Within the recent explosion of literature on metaphysical explanation, it is orthodox to suppose that the core function of metaphysical explanation discourse is to track metaphysically robust relations (such as relations of ground, or other dependence relations), and hence that the core function of these judgements is to assert propositions that have substantial truth-conditions (i.e. truth-conditions that outstrip the T-schema). We call such views *cognitivist descriptivist* views. They are cognitivist because they say that the core function of judgements of ⌜x *because* y⌝ is to assert propositions (and hence to report truth-apt beliefs) rather than, say, to express some non-cognitive attitude. They are descriptivist because they say that the core function of the discourse in which these propositions are asserted is to accurately describe some substantive feature of the world: in particular, to assert truths about the obtaining of metaphysically robust relations.

Within this research program it has often been argued that instances of ⌜x *because* y⌝ are true iff there is a pair of facts, [x][[3]](#footnote-3) and [y], such that [x] obtains iff the sentence x, is true, and [y] obtains iff sentence y is true, and each fact is in some good sense what the corresponding sentence is *about*, and a particular explanatory connection obtains between [y] and [x].[[4]](#footnote-4) In turn, it is often thought that *grounding* is the explanatory connection that must obtain between [y] and [x] in order for ⌜x *because* y⌝ to be true.[[5]](#footnote-5) So, for instance, ‘the Jenga tower exists *because* the Jenga blocks are arranged thus-and-so’ is true just in case [the Jenga blocks are arranged thus-and-so] grounds [the Jenga tower exists]: the former fact bears the relation of ground to the latter fact.

Against this backdrop, the view we introduce in this paper is quite radical, for it is both non-cognitivist and non-descriptivist.[[6]](#footnote-6) We will make the case that the core function of the discourse in which we find judgements of ⌜x *because* y⌝ is not to accurately describe or track metaphysically robust relations, and the core function of judgements of ⌜x *because* y⌝ is not to assert propositions (and hence to report truth-apt beliefs) that are made true by the presence of metaphysically robust relations. Rather, the core function of the discourse in which such judgements are embedded is to express attitudes of commitment, at a context[[7]](#footnote-7) C, to certain norms governing interventional conduct, and to recommend to those at contexts relevantly similar to C, that they commit to those norms.

Nevertheless, we say, while the core function of judgements of ⌜x *because* y⌝ is to *express* these attitudes, instances of ⌜x *because* y⌝ have perfectly ordinary literal semantic content (content that is about [x] and [y] and not about anyone’s non-cognitive attitudes) and instances of ⌜x *because* y⌝ are often true in a minimal or deflationary sense. That is, such propositions are often true, but are not made true by metaphysically robust relations.

We argue that this version of non-cognitivist non-descriptivism provides a pleasingly nuanced account of what we are doing when we make such judgements.

Here is the plan. §2 introduces the non-cognitivist component of our account: a kind of norm expressivism according to which the core function of judgements of ⌜x *because* y⌝ is to express commitment to and recommendation of, certain interventional norms. In §3 we introduce the non-descriptivist component of our view, according to which propositions of the relevant form have ordinary literal semantic content, and sometimes in utterances such propositions we assert this literal semantic content, which can be minimally true. §4 and 5 argues that non-cognitivism and non-descriptivism together provide a pleasing account of our ordinary practises, and of the context-sensitivity of our judgements.

**2. Norm Expressivism**

We are taking metaphysical explanation discourse to consist in our various judgements, and utterances, of the form ⌜x *because* y⌝. Of course, philosophers engaged in discussion of metaphysical explanation are often interested in the nature of the phenomenon itself, rather than in the status of particular judgements of this form. Much of this philosophical discourse is discourse *about* metaphysical explanation discourse, rather than being metaphysical explanation discourse itself. By offering a non-cognitivist account of metaphysical explanation we add to this discourse about the discourse.

Before we get to the nitty gritty of our account, a few preliminaries are in order. First, we focus on an account of the *core* function of the first-order discourse. We are interested in what people are, ordinarily, doing when they judge that ⌜x *because* y⌝. When philosophers are engaged in their meta-level discourse they often *mention* judgements of ⌜x *because* y⌝ rather than making those judgements. We are not offering an account of what philosophers are doing when mention those judgements, only when they make them.

Second, our view is a view about the core function of the discourse in which we find judgements such as of ⌜x *because* y⌝: namely that the core function of this discourse, and, in turn, of judgements of the form ⌜x *because* y⌝ is to express certain attitudes. This does not commit us to the view that *every* such judgement of ⌜x *because* y⌝ expresses such an attitude. Just as one can be a moral non-cognitivist (and think that the core function of moral judgements is to express certain attitudes), and yet allow that sometimes we make such judgements without expressing those attitudes, so too, one can think that while that the core function of judgements about metaphysical explanation is to express certain attitudes, it is nevertheless the case that sometimes these judgements do not express those attitudes.

Our approach, here, follows Price’s broadly pragmatist approach. Price (2001:105) notes that “a philosophical account of a problematic notion—that of causation itself, for example—needs to begin by playing close attention to the role of the concept concerned in the *practice* of the creatures who use it….we cannot hope to explain this anthropological fact if we begin…by looking for something for causation to *be*…Instead…we need…to ask what role such notions play in the lives of the creatures concerned—why creatures like us should have come to describe the world in these causal terms.”

We seek to provide an account of what role judgments of the form ⌜x *because* y⌝, play in everyday interactions between agents. We focus on making sense of this discourse as a whole, and on judgements that we take to be commonplace. We are not alone in taking such judgments to be commonplace. For instance, seminal work by Schaffer (2009:375) makes the case that this kind of explanation is “a natural and intuitive notion, for which there exist clear examples, and clear formal constraints.” He mentions that when his first-year students, who do not (yet) view the world through a peculiarly metaphysical lens, come across the notion of metaphysical explanation in the philosophy classroom, it is a notion with which they are already acquainted. Dasgupta (2017:74-76) similarly argues that metaphysical explanation is “intuitive and familiar, and at the same time useful in framing a number of philosophical debates” and that “examples are ubiquitous. Why is a faculty meeting occurring? Because the faculty are gathered in a room discussing matters of importance to the department, etc. Why is this water hot? Because its mean kinetic energy is high. Why have I lost this game of chess? Because my king is in check-mate.” Indeed, it is “*an everyday concept used by the masses*. When I explain the concept to non-philosophers they recognize it immediately and talk intelligibly about it, offering examples of [metaphysical] explanations in their own fields of biology, economics, journalism, or cooking. To them it is not a new concept.”

Likewise, Glazier (2020:121) notes that “[f]or its enthusiasts, [metaphysical] explanation is both ubiquitous in ordinary life and central to many of philosophy’s biggest questions”. Trogdon’s (2013) influential overview of early(ish) debates about grounding usefully coins the term *quotidian*, for the common view that metaphysical explanation is part of ordinary, everyday thinking.

Indeed, we think most metaphysical explanation discourse takes place outside the academic context. That is why when investigating the core function of judgements of the form ⌜x *because* y⌝ we ought not be focussing purely on philosophical judgements, but, more broadly, on everyday judgements of that form.

Our view about the core function of judgements of ⌜x *because* y⌝ is a version of norm expressivism. It is the view that such judgements express attitudes of commitment to, and recommendation of the acceptance of, certain norms governing what we call *interventional* *conduct*. In what follows we motivate the idea that there is some tight connection between metaphysical explanation discourse and interventions, and explain what we take interventions to consist in. We then explicate our account of norm expressivism in terms of interventional conduct.

*2.1 Interventions and Explanation*

There is an intimate relationship between causal explanations, the imagining of interventions and the representation of interventional affordances (that is, the representation that intervening on one part of the world in a certain way will thereby intervene on some other part of the world in a certain way). There is ample evidence that imagining interventions is vital in causal reasoning[[8]](#footnote-8) and in causal explanation.[[9]](#footnote-9) Our practice of causally explaining one thing in terms of another is crucially connected to these interventional representations since it is only by appealing to interventions, both imagined and actual (more on this shortly) that we can ascertain both which pairs of events or facts are such that a causal explanation obtains between them, and, further, which direction the explanation runs in (i.e., which is the explanans and which the explanandum).[[10]](#footnote-10) Hence representing the interventional affordances in one’s environment is key to effective agency: doing so tells us what things are levers by which we can manipulate other things.

For this reason, interventions are not only the focus of recent philosophical work on causation and causal explanation,[[11]](#footnote-11) but have been widely studied in the psychological literature. There is substantial evidence that individuals like us represent a range of counterfactual interventions: that is, we represent that *if* we intervened on some part of the world (though we in fact do not) we would thereby intervene on some other part of the world.[[12]](#footnote-12) For instance, when people are asked what will happen if an experimenter intervenes in the world in a particular way, there is good evidence that participants imagine the intervention and its consequences.[[13]](#footnote-13) Most empirical work on this phenomenon has centred around these imagined counterfactual interventions: interventions that people could (in some sense of could) perform, but instead of actually doing so they merely imagine the counterfactual scenario in which they do, and simulate what would happen if they did.

That the representation of interventional affordances is intimately connected with causal explanation strongly suggests that they are also connected with metaphysical explanation. Indeed, the idea that there is such a connection is not new. Miller & Norton (2017) make this case, while and Schaffer (2016) and Wilson (2018a; 2018b) have independently developed accounts of grounding that make use of the same formalism as do interventionist accounts of causation.[[14]](#footnote-14) Their idea is that by developing models that represent which facts are interventional affordances for which, we can thus establish both what causes what, and what grounds what. While this is not an avenue we pursue, we think that people are inclined to judge that there is a metaphysical explanation present when they are inclined to judge that there is an interventional affordance present.

For example, we imagine intervening on some part of the chair in order to intervene on the chair; we imagine intervening on the brain in order to intervene on the mind; we imagine intervening on Gibbard in order to intervene on {Gibbard}; we imagine intervening on categorical properties to intervene on dispositional properties; we imagine intervening on truthmakers in order to intervene on what is true; and we imagine intervening on determinate properties in order to intervene on determinable properties. In each of these cases, it seems, the intervention that we imagine tracks our judgements about what metaphysically explains what.

While humans have reached a stage at which we sometimes want to know things, and understand things, just for the sake of knowledge and understanding, our desire to understand the world surely stems from evolutionary pressure for fitness. Creatures that are better able to understand the world around them in such a way that they can manipulate it to attain food, shelter, clothing, warmth, safety, mates, and so on, do better than those who cannot. There’s a reason that tool-using creatures are usually pretty smart: these are creatures that have learned new ways to manipulate the world to their advantage.

Causal reasoning is important because it allows us to work out which ways of intervening on the way the world is now, bring about alternations in the way the world is later. These sorts of interventions were, and are, crucial to our survival. The function of causal explanation is, in the end, to work out how best to intervene in the world now, in order to bring about certain desirable effects. We submit that metaphysical reasoning is important for the same reason: because it allows us to work out which ways of intervening on one part of the world allow us to intervene on some other part of the world, albeit non-causally. Likewise, the function of metaphysical explanation discourse is ultimately to work out how best to intervene in the world, in order to non-causally bring about certain desirable effects.

In ordinary contexts when individuals make claims about what metaphysically explains what, they are interested in ordinary matters such as heating water, or making moral choices, or building bicycles, or curing headaches. They want to make plans to intervene in the world in ways that best allows them to fulfil their goals. So, we say, judgements of ⌜x *because* y⌝ at contexts are expressions of commitment, at that context, to going about engaging with the world in a certain way. They express the individual’s practical intentions at that context, and in expressing those practical intentions the individual says to others: if you are at a context like this, then this is a good way to go about things. That is, these judgements express not only commitment to, but recommendation of, these norms. We call these *interventional norms*.

*2.2 Interventional Norms*

What, then, is an interventional norm? We begin by explicating the general idea of commitment to, and recommendation of, norms, before explaining what it is to commit to (and recommend) interventional norms.

Today I have plans for what I will do tomorrow. Some things I *could* do tomorrow, I have ruled out as options. I have ruled out the option of stealing and eating my friend’s lunch, even though that’s a thing I could do. In fact, I have ruled out a whole bunch of behaviours, and ruled in others. According to Gibbard (1990; 2003) when I judge that it is wrong to steal and eat someone’s lunch, I am expressing a commitment to a certain norm: I am expressing that I have ruled out plans that include going to work and stealing and eating someone’s lunch.[[15]](#footnote-15)

That, however, cannot be the full story. Suppose I am alone on a deserted island on which there are no cats. It seems that the option of eating some cats to stay alive is not available to me. So I cannot plan to refrain from eating cats. That poses a problem if we think that while on the island, I can nevertheless judge that eating cats is wrong. The solution is to see that I can commit to *conditional plans*. I can ask myself: *if there were cats on this island, would I eat them?* If my answer is that I have a conditional plan to abstain from eating cats on this island, even if there were some cats to eat, then I have a conditional plan to avoid cat-eating. By ruling out (and ruling in) certain options by my plans and conditional plans, I commit to various norms of conduct.

One commits to an *interventional* norm when one rules in (and out) certain interventions. One rules in an intervention if one has plans to engage in that intervention, and rules out an intervention if one excludes any such plan. What are these plans? They are plans to intervene on some part of the world by intervening on some other part of the world. More carefully, interventional plans are conditional plans: plans that are conditional on our goals and desires. My plan to intervene on the mean molecular energy of the water in order to intervene on its temperature, say, is conditional on my wanting to intervene on the temperature of the water. So one rules in an intervention if one has a conditional plan to engage in that intervention given certain goals, and rules out an intervention if one has no such conditional plan.

The conditional interventional plans to which individuals are committed vary depending on context. That’s because at different contexts different interventions are salient, practical, and relevant. At Jemima’s context she does not represent that intervening on the mass of the Earth is a way to intervene on its gravity, because she no capacity to intervene on the Earth’s mass, and that interventional affordance is simply not salient to her. So even if, in fact, Jemima desires to intervene on the Earth’s gravity, at that context she has no conditional plan to do so by intervening on the Earth’s mass. Perhaps, though, there are contexts Jemima could come to occupy at which that affordance is salient. If Jemima develops super-human powers by, say, eating a radioactive slug, then she might represent there to be such an affordance, and she might commit to a conditional plan to intervene on the Earth’s mass in order to intervene on its gravity.

Quite generally, then, individuals commit to different conditional interventional plans at different contexts, and hence commit to different interventional norms at different contexts. So in what follows we talk about individuals committing to interventional norms *at contexts.*

Then according to the version of non-cognitivism we are offering, the core function of a judgement of ⌜x *because* y⌝, made at context C, is to express commitment to (and recommendation of the acceptance of; more on this shortly) a certain interventional norm: namely the norm of ruling in the plan of intervening on [y] in order to intervene on [x], conditional on wanting to intervene on [x]. So for example when Jemima walks into the bicycle shop and judges that ‘the bicycle exists *because* the parts are arranged bicycle-wise’ she is expressing (*inter alia*) a commitment, *at that context*, to an interventional norm that rules in intervening on the arrangement of bicycle-parts in order to intervene on whether there is a bicycle, conditional on wanting to intervene on whether there is bicycle.

Jemima also, we suggest, expresses a recommendation to others to adopt this interventional norm, conditional on them occupying a context that is relevantly similar to that occupied by Jemima herself.

There is good evidence that we recommend to others that they intervene in the world in certain ways, at certain contexts, in order to achieve certain goals. This kind of recommendation is a staple of human interaction. Contemporary psychological research has focussed on the age at which we begin to make[[16]](#footnote-16) and respond to such recommendations[[17]](#footnote-17) as well as the extent to which other species engage in this behaviour.[[18]](#footnote-18)

We take it to be plausible that when individuals commit to certain interventional norms at contexts, they are, at least tacitly, recommending these norms. After all, if the individuals in question did not think these appropriate norms in the relevant context, then it is hard to see why they would commit to them. Since often, committing to such norms involves modelling them for others, it seems reasonable to suppose that these individuals are recommending those norms to those situated in similar contexts.

*2.3 Apt Non-cognitive States*

So far all we’ve said is that the core function of a judgement of ⌜x *because* y⌝ is to express certain non-cognitive attitudes. Since non-cognitive attitudes are not truth-apt there can be no question of whether we express something true when we express those attitudes. Pretty clearly though, some interventional norms are better than others. That’s because, again pretty clearly, some interventions work, and some do not. Painting the parts of the chair is a good way to intervene on its colour, doing a colour-dance is not; rearranging our social institutions is good a way to intervene on whether our society is just, talking to a praying mantis is not. And so it goes.

To capture the fact that some interventions will have the desired effect and others will not, we want to be able to say that some judgements of ⌜x *because* y⌝ are *apt*, and others are not. A natural thought, at this point, is that what makes such judgements apt is the presence of certain dependence relations in the world. A judgement of ⌜x *because* y⌝ is apt, just in case [x] depends on [y] (where [y]’s grounding [x] would be a specific case in point). If we were to take this route, however, it would be unclear why we shouldn’t think that the core function of this discourse is really just to point out the presence of these dependence relations. That, in turn, would suggest that the core function of judgements ⌜x *because* y⌝ is to assert propositions that are made true by metaphysically heavyweight dependence relations. That would firmly lead us back towards a descriptivist and cognitivist account of the phenomenon.

Given the benefits of non-cognitivist non-descriptivism—to which we will return—we are resistant to this line of reasoning. Instead, we say that a judgement of ⌜x *because* y⌝ is apt, just when there obtains the appropriate *interventional fact*. As we conceive of them, interventional facts (no capitals) are not metaphysically robust. These are the sorts of deflationary facts that we get for free according to certain deflationary meta-ontological views such as those of Schiffer (2003) and Thomasson (2015). On such views we can make trivial inferences from obviously true claims such as ‘the labradoodle is black’ to ‘the labradoodle has the property of being black’, to ‘there are properties’. But to be committed to a property is only to be committed to a metaphysically lightweight thing, since to say that there are properties is to say no more than that, *inter alia,* labradoodles are black. Likewise, we can move from the obvious truth (say) that ‘intervening on y is a way to intervene on x’ to infer that ‘it is a fact that intervening on y is a way to intervene on x’ to ‘there are interventional facts.’ Importantly, there being an interventional fact, in this lightweight sense, does not imply that there are metaphysically heavyweight dependence relations.

In particular, we think that there can be an interventional fact of this deflationary kind where there is no attendant dependence relation (even assuming such relations exist), and conversely, there can be a dependence relation where no interventional fact obtains.

To see the former, suppose that in our world there are dependence relations, but that [x] does not depend on [y]. Nevertheless, as a highly local, context-dependent matter of fact, at contexts of a certain sort, including Freddie’s context, intervening on [y] is a way to intervene on [x]. How can that be? Well even if you think that the only way there could be reliable interventions on [x] via [y], is if [x] depends on [y], one surely should grant that there could be unreliable interventions that are not like this: interventions that succeed in particular contexts for certain local contingent reasons. This case is one such case. While in general intervening on [y] is not a reliable way to intervene on [x], *at Freddie’s context* he can (reliably) intervene on x by intervening on y. So it is a fact that at that context intervening on [y] is a way to intervene on [x], despite its not being the case that there is any dependence of [x] on [y].

To see how this might go, suppose that there is a classroom full of bald people, and Andrew is standing outside the classroom door with a full head of hair. The fact that everyone in the classroom is bald depends (we shall suppose) upon all the bald people in the classroom, and not (we shall suppose) on any facts about Andrew. And in general, it’s not the case that one can reliably intervene on whether everyone in a given room is bald by intervening on Andrew’s location. Nonetheless, in this case there is an interventional fact: by intervening on which side of the threshold he stands on, Andrew can intervene on whether everyone in the classroom is bald. So at Andrew’s context, he can reliably use his own location as a means by which to intervene on whether everyone in the classroom is bald. This is so, to reiterate, despite the fact that everyone in the classroom is bald *not* non-causally depending on any facts about Andrew.

Now let’s consider a case in which there is a dependence relation present, but there is no interventional fact. This will be the case when one fact depends on, and is overdetermined by, multiple facts. For example, suppose that [there is at least one expressivist] depends on [Gibbard is an expressivist]. While in general, when one fact depends on another there are at least some interventions on the latter that will intervene on the former, that’s not so in this case, because of the existence of other expressivists. So it’s not the case that intervening on [Gibbard is an expressivist]—by convincing him of the merits of moral cognitivism, say—is a way of intervening on [there is at least one expressivist]. To be sure, if we imagine a counterfactual where we hold fixed that there are no other expressivists, *then* intervening on [Gibbard is an expressivist] is a way of intervening on [there is at least one expressivist]. But that’s not sufficient for it to be a fact that intervening on [Gibbard is an expressivist] is a way of intervening on [there is at least one expressivist] at contexts in our world. More generally, it can be the case that [x] depends on [y], and yet fail to be the case that intervening on [y] is a way of intervening on [x].[[19]](#footnote-19)

On our view, then, judgements of ⌜x *because* y⌝ are apt only when there is the appropriate interventional fact regarding [y] and [x]. That, however, does not commit us to thinking that there are any heavyweight dependence relations between these facts, nor, if there are, that these are what render such judgements apt. So we have no reason to suspect that really, the purpose of these judgements is to track such metaphysically heavyweight relations. That brings us to the second aspect of view: non-descriptivism.

**3. Non-descriptivism and minimal truth**

The second aspect of our view is a kind of non-descriptivism. While our view is that the core function of judgements of ⌜x *because* y⌝ is to express certain non-cognitive attitudes, we do not claim that the literal semantic content of ⌜x *because* y⌝ is to report a non-cognitive state. Rather, we think that ⌜x *because* y⌝ has a perfectly ordinary literal semantic content that is about the world: it is about [x], and [y], and some explanatory connection between them. So we allow that utterances of ⌜x *because* y⌝ often assert the literal semantic content of these propositions. Further, we allow that this semantic content can be minimally true, that is, true according to a minimal or deflationary theory of truth. So some particular instance, of ⌜x *because* y⌝, say h *because* k, is true iff h *because* k. Hence the core function of an utterance of ‘h *because* k’ is to express certain non-cognitive attitudes, but nevertheless that utterance (probably) also asserts the literal semantic content of the proposition, and what is asserted will be minimally true iff x *because* k. Nothing metaphysically heavyweight is required for the truth of such propositions.

What we deny, then, is that such assertions have substantial truth-conditions. If the purpose of such judgements were to track robust metaphysical relations, such as for instance to track dependence relations, then we would expect the core function of judgements of ⌜x *because* y⌝ to be to assert propositions that have robust truth-conditions: namely, propositions that are true just in case the relevant dependence relations obtains. This is precisely what we do not think is going on. Indeed, we think that the fact that interventional facts can come apart from dependence relations gives us reason to think that the purpose of these judgements is not to track dependence relations. For consider the case of Freddie. It seems to us that what matters to Freddie is how he can use some parts of the world to intervene on other parts. So what matters to Freddie are interventional facts. Or, perhaps better, what matters to Freddie are the interventional facts that obtain at his context. That gives us reason to think that the purpose of Freddie’s judgements is not to track dependence relations.

There are a few notable features of this aspect of the account.

First, it allows the account to avoid Frege/Geach embedding problems. Since we accept that propositions of the form ⌜x *because* y⌝ have literal semantic content, they can be embedded in the usual manner. So for instance, the following argument is straightforwardly valid, because ‘x *because* y’ is used throughout with the same perfectly ordinary semantic content.

1. If x *because* y, then we should all do the polka.
2. x *because* y.
3. Therefore, we should all do the polka.

Second, by appealing to this aspect of the account we can accommodate there being judgements of the form ⌜x *because* y⌝ that do not appear to express commitment to, and recommendation of, interventional norms. The cases we have in mind are those in which no intervention can be performed. For instance, Sally might judge that salt dissolves in water *because* water is H2O (rather than XYZ) or that Tim the time-traveller didn’t succeed at killing his grandfather *because* that would have prevented him from being born,[[20]](#footnote-20) or that Cicadas have the life-spans they do *because* 13 and 17 are prime numbers, and so on. Moreover, philosophers, at least, tend to judge that {2} exists *because* 2 exists. Assuming that 2 and {2} exist necessarily, then no one can intervene on either. So no judgement that {2} exists *because* 2 exists can express a commitment to, or recommendation of, an interventional norm regarding {2} and 2. Quite generally these cases raise problems for our account, since one cannot commit to a plan to do P, if one believes that P is impossible.

To see what’s going on with these cases it’s important to remember that our view is not that all judgements of this form express commitments to, and recommendations of, interventional norms. Rather, our claim is that expressing such commitments and recommendation is the core function of making such judgments. That is, it’s the *reason why* we have a discourse in which such judgments feature. That is entirely consistent with us sometimes making such judgements without expressing these non-cognitive attitudes.[[21]](#footnote-21)

Recall that on our view the propositions in question have literal semantic content that can be minimally true. Sometimes our judgements *only* assert that semantic content. This will be the case when we make such judgements and there are no possible interventions we could make, and hence no interventional norms we could reasonably be committed to, or recommend. These cases are in some good sense parasitic on cases in which interventions can be performed: on cases in which our judgements do express the relevant non-cognitive attitudes. They are parasitic just in the sense that, on our view, we wouldn’t be in the business of making such judgements if they didn’t, usually, express commitment to interventional norms. But that doesn't mean there is anything odd, or peculiar about such judgements, or that they present themselves to us as being in some way parasitic on other kinds of judgments. We don’t think that they do.

Why, though, do we make such judgements? And why do we make the particular judgements that we do? We think the answer to these questions is, roughly, that in these cases we use the same sort of cognitive machinery that we use in ordinary contexts to judge that ⌜x *because* y⌝. In particular, we suggest that there is cognitive machinery that allows us to make these judgements by enabling us to determine, at contexts, whether there is an interventional fact. Even where there is no such fact, however, that machinery can be, as it were, run off-line in a kind of simulation, where the output of that simulation guides our judgements regarding whether ⌜x *because* y⌝. The general point is that the cognitive machinery can function perfectly effectively even in the absence of any interventional fact, and it is the functioning of this machinery that explains why we make such judgements even when they do not express commitment to interventional norms. We will have more to say about this machinery in §5. Before doing so, however, we want to turn to consider an upshot of our view: namely the context-sensitivity of judgements of ⌜x *because* y⌝. We argue that this is a welcome consequence of our view, and that it nicely accommodates the sorts of judgements that we in fact find.

**4. Context-sensitivity**

Given that, on our view, the core function of judgements of ⌜x *because* y⌝ is to express commitment to, and recommendation of, certain interventional norms, we would expect these judgements to be context-sensitive. That is, we would expect people to make different judgements in different contexts. For instance, consider *‘Stressed* *Bert’*, a man who is stressed and has high cortisol levels. Let’s suppose that while Bert’s being stressed (his mental state) and his high cortisol levels (his brain state), co-vary (he has one just in case he has the other) the former is distinct from the latter. Bert has two friends, Fred and Maria. Maria is a pharmacist, and Fred is a therapist. Maria is able to dispense drugs that will intervene on Bert’s brain states so as to reduce his cortisol levels, thereby also reducing his stress levels. Fred is a very good therapist, who is able to teach Bert to directly intervene on his stress levels in various ways, thereby also reducing his cortisol levels. Let’s suppose that Bert has, at this time, done neither of these things.

Tomorrow he makes two appointments. In the morning he goes and visits Fred. Fred explains to him how to go about intervening on his stress levels. At the end of their appointment Fred says to Bert ‘you have high cortisol levels *because* you are stressed’. It seems to both Bert and Fred that what Fred says is an explanation. In the afternoon Bert goes to visit Maria. Maria fills a prescription for a new drug, and explains to Bert how the drug works to lower his cortisol levels. At the end of their discussion, Maria says to Bert ‘you are stressed *because* you have high cortisol levels’. It seems to both Bert and Maria that what Maria says is an explanation.

We take there to be nothing particularly odd about the case just described. Moreover, we take it to be natural to conclude that in one context we will judge that Bert has high cortisol levels *because* he is stressed and in the other we will judge that Bert is stressed *because* he has high cortisol levels. That is, our judgements are context-sensitive. Recent empirical evidence supports this idea (Latham & Miller MSa).

While philosophers shown surprisingly little interest in understanding or accommodating this apparent context-sensitivity, a number of proposals have recently been put forward. These proposals share one thing in common: they all attempt to accommodate the context-sensitivity of judgements by holding that metaphysical explanation itself is context-sensitive. That is, they hold either that (i) particular instances of ⌜x *because* y⌝ take different truth-values at different contexts, so that, for example ‘Bert has high cortisol levels *because* he is stressed’ is true at some contexts and false at others, and likewise for ‘Bert is stressed *because* he has high cortisol levels’ or (ii) particular instances of ⌜x *because* y⌝ are metaphysical explanations (or, perhaps, are *successful* metaphysical explanations) at some contexts but not others.

Thompson (2016), for instance, argues that “[w]hat makes for successful metaphysical explanation will depend (to an extent) on features of agents” (p. 398) and in her (2019) she develops a context-sensitive account of metaphysical explanation whereby metaphysical explanations are answers to a certain kind of why-question: a ‘what-makes-it-the-case-that’ question. Whether or not an utterance counts as a metaphysical explanation, for Thompson, depends on “epistemic constraints imposed by the context in which a relevant question is asked” (p.98). For Thompson, whether something counts as a metaphysical explanation *for an individual at a context*, depends in part on the goals and capacities of that individual, and in part on the obtaining of the relevant dependence relation.

Dasgupta (2017) has also defended what he calls *anti-realism* about metaphysical explanation: the view that what metaphysically explains what is relative to our interests and concerns, which in turn “may vary from culture to culture or time to time” (p. 78). According to Dasgupta’s preferred anti-realist picture “two cultures might offer conflicting [metaphysical] explanations and yet there may be no fact of the matter who is “really correct”: each explanation may be correct relative to their respective interests and concerns.” (p. 89). Precisely what determines which metaphysical explanation is correct relative to a culture is not something Dasgupta aims to answer.

A rather different approach is advocated by Norton & Miller (2019), who defend what they call a ‘psychologistic theory of metaphysical explanation’. According to this view, an instance of ⌜x *because* y⌝ is true, for an individual just in case ⌜y⌝ is true, ⌜y⌝ entails ⌜x⌝ and the *community in which that individual is embedded* is disposed to have certain beliefs about what metaphysically explains what, after engaging in a complicated process of group deliberation which Norton & Miller call THE PROCEDURE.

Non-cognitivist non-descriptivism can make good sense of the idea that such judgements are context-sensitive, without being committed to holding that the truth-value of instances of ⌜x *because* y⌝, or whether such instances *are metaphysical* explanations, varies across contexts. It is natural to say that these judgements vary across contexts simply because different people commit to, and recommend, different interventional norms at different contexts. Indeed, out view predicts that individuals will express commitment to those interventional norms, which, at a context, are relevant, salient, and practically useful, because we know from empirical evidence on causal explanation that these are the sorts of factors that determine whether people represent and cite particular interventions when proffering causal explanations.

For instance, we know that a good many interventions are not represented at all, even when the relevant judgement would be apt, in the sense articulated in §2.3. Research shows that from amongst the various interventions one could perform that would achieve a certain goal, people are more likely to represent that intervention which is in some way salient than those interventions that are not.[[22]](#footnote-22) There is also evidence that amongst all the interventions which would achieve our goals, we are much more likely to countenance interventions that we in fact have the power to bring about, and that expectations about how scenarios tend to play out inform our predictions about which interventional affordances obtain, and that, in turn, constrains which interventions we end up representing.[[23]](#footnote-23)

So we would expect individuals to commit to, and recommend acceptance of, interventional norms, at a context, which are relevant to their goals, which are salient as a way of serving those goals, and which are practical: interventions that the individual can perform at that context. Since individuals’ capacities vary across contexts, and since which interventions are salient also varies, we would expect individuals to commit to, and recommend, somewhat different norms at different contexts.

So, we say, Maria’s and Fred’s different judgements reflect and express the different interventional norms to which each is committed at their respective contexts. Fred’s judgement that Bert has high cortisol *because* he is stressed expresses his commitment to, and recommendation of, intervening on Bert’s mental state to intervene on his brain state, while Maria’s judgement that Bert is stressed *because* has high cortisol levels expresses her commitment to, and recommendation of, intervening on Bert’s brain states to intervene on his stress levels. Fred and Maria are committed to interventional norms that are salient, relevant, and practical at their contexts.

Indeed, our view predicts that in the Stressed Bert case, there will be contexts at which people will make *both* sorts of judgements. This will be so when people commit to and recommend both norms. Perhaps, having seen both Fred and Maria and thus having come to represent the interventions each recommends, Bert will commit to both norms, and thus come to judge both that he is stressed *because* he has high cortisol levels and that he has high cortisol levels *because* he is stressed. And indeed, Latham & Miller (MSb) found some evidence that people make symmetrical judgements in the case of Stressed Bert.

It need not, however, follow from this that instances of ⌜x *because* y⌝ are true at some contexts and false at others, or *are explanations* at some contexts but not others. Recall that we are offering a deflationary view about the truth of the proposition expressed by ⌜x *because* y⌝: ⌜x *because* y⌝ is true iff ⌜x *because* y⌝. Then one might think that ⌜x *because* y⌝ is true, or false, *simpliciter*, (i.e., takes the same truth-value across all contexts) regardless of the interventional norms to which individuals are committed/recommend. After all, on our view the truth of these propositions is lightweight, so there don’t have to be special facts that obtain at some, but not other, contexts, and which make the proposition true at some, but not other, contexts.

Alternatively, one might hold that whether ⌜x *because* y⌝ is true at some context bears a tight connection to whether a judgement of ⌜x *because* y⌝ at that context is apt. That is, one might think that one way for it to be that ⌜x *because* y⌝ is true at a context, is for it to be that we judge ⌜x *because* y⌝ at that context, *and* the judgement is apt: it expresses a fitting attitude. If so, one will think that the truth of instances of ⌜x *because* y⌝ is indeed itself context-sensitive.

On this latter view ⌜x *because* y⌝ is never true or false *simpliciter*, but is always true or false relative to an individual at a context. We are tempted to think that insofar as one is drawn to the idea that not only are our judgements of ⌜x *because* y⌝ context-sensitive (and rightly so) but also that the truth-value of ⌜x *because* y⌝ is itself context-sensitive, that our version of this context-sensitivity is simpler than the cognitivist alternatives. Those cognitivist views employ substantial truth-conditions for instances of ⌜x *because* y⌝. So views such as that of Dasgupta, Thompson, and Norton & Miller all require fairly complicated machinery in which the truth of an instance of ⌜x *because* y⌝—or whether it is an explanation—for an individual occupying a certain context depends in some way on the psychology of the individual, and perhaps the dispositions of the community in which she is embedded, as well as the presence of some metaphysically heavy-weight relation.

The second version of our view is somewhat simpler than these context-sensitive cognitivist accounts. According to this second version, ⌜x *because* y⌝ is true at a context just in case the individual at that context judges that ⌜x *because* y⌝, and the judgement is apt because there is an interventional fact regarding [y] and [x]. There is no need to appeal to the individual’s cognitive states or those of her community.

**5. Explaining Ordinary Practices**

There is a range of ordinary practices that our account can neatly accommodate and explain. We focus on three of these: disagreement, agreement, and judgements in contexts in which no intervention is possible. We take these in order.

Consider, first, disagreement. Since on our view utterances of ⌜x *because* y⌝ do assert (minimally) truth-apt propositions, we can understand disagreement between parties in a perfectly standard manner in terms of the assertion of propositions with incompatible content (such as ⌜x *because* y⌝ and ⌜¬(x *because* y)⌝). But our account is much richer and more interesting than that. When individuals occupy very similar contexts, and one judges ⌜x *because* y⌝ and one judges that ⌜¬(x *because* y)⌝, in ordinary circumstances where these judgement express non-cognitive attitudes, these individuals are expressing commitment to, and recommendation of the acceptance of, *incompatible* interventional norms. One party recommends adopting one set of interventional norms at a context, and another recommends a different set of interventional norms *at a relevantly similar context.* Since at most one recommendation can be adopted, there is *another* perfectly good sense in which these individuals are disagreeing: they are providing incompatible recommendations. Call this *attitudinal disagreement.*

That there are sometimes cases of attitudinal disagreement is entirely consistent with there being a range of cases in which we want to say that there is no attitudinal disagreement at all. If one individual judges that ⌜x *because* y⌝ and one judges that ⌜¬(x *because* y)⌝, and each of these individuals occupies a very different context, then they are not making incompatible recommendations. One is recommending certain norms in one context, and one is recommending different norms at a quite different context. No attitudinal disagreement obtains.

Recall too, that we have said that sometimes judgements of ⌜x *because* y⌝ at a context are apt, and sometimes not. They are apt when at the relevant context, the relevant interventional fact obtains. That raises the question of how we go about determining whether such a judgement is apt. We suggest that we are able to make determinations about whether a judgement in some context is apt, by simulating the interventional norms to which we *would* be committed, *were* we to occupy that context.

This opens up further scope for thinking about different kinds of attitudinal agreement and disagreement. Suppose that Jemima is evaluating Melanie’s judgement of ⌜x *because* y⌝. Jemima will (and should) take that judgement to be apt, just in case when Jemima simulates being in Melanie’s context (with Melanie’s capacities and local environment) she sees that at that context she too would commit to the interventional norm to which Melanie has expressed commitment: namely the norm of intervening on [y] in order to intervene on [x]. Judgements of aptness, then, are judgements that the relevant judgement being made at some other context expresses the norm to which one *would* commit, *were* one at that context.

Thus, it might be that although Melanie judges that ⌜x *because* y⌝ and Jemima judges that ⌜¬(x *because* y)⌝, each judges that were they in the context occupied by the other, they too would be committed to the interventional norms to which the other is committed. Then each judges that the other’s judgement is apt. Thus there is no attitudinal disagreement between them.

On the other hand, suppose each judges that even were they in the context occupied by the other, they would nevertheless *not* commit to the interventional norms to which the other is committed. Then each judges that the other’s judgement is not apt, and so they disagree about which set of interventional norms to commit to, and recommend, *at each of the two contexts.* In this case there is an attitudinal disagreement, since each takes the other to be committing to the wrong interventional norm, even given the context she occupies.

So our account offers a certain kind of richness in understanding and classifying different kinds of disagreement. We can focus both on whether parties assert incompatible propositions, and we can also focus on whether parties make incompatible recommendations. We can, then, capture two senses in which parties can be disagreeing (or not): attitudinal disagreement and propositional disagreement. We take this to be a feature of the view.

Second: agreement. Subjects like us often make the same judgments about what metaphysically explains what. Our view neatly accommodates this fact.

Consider the rather recherché case that philosophers often discuss, regarding objects and the singleton sets containing them. Philosophers seem universally to judge that a given set exists *because* its member(s) exist (at least, that’s so amongst philosophers who have expressed their judgements about this case). Consider a particular object and singleton set: Gibbard (the expressivist), and {Gibbard} (the singleton). We think that when philosophers judge that {Gibbard} exists *because* Gibbard exists, (as opposed to simply mentioning that judgement) they express commitment to certain interventional norms. Moreover, we can explain why we find such convergence regarding the norms to which we commit (and convergence in judging that {Gibbard} exists *because* Gibbard exists).

We noted previously that people are inclined to represent those interventions that are relevant, salient, and practical. While many philosophers have no opportunity to intervene on Gibbard (we are writing this under pandemic-induced conditions of physical distancing that make this abundantly clear) it is nevertheless the case that intervening on Gibbard is considerably more practical than intervening on {Gibbard}. Insofar as one wants to intervene on either of these, and hence has a conditional plan, individuals like us will almost certainly plan to intervene on Gibbard, a concrete object on which we have the power to intervene, in order to intervene on {Gibbard}, an abstract object on which we do not have the power to intervene. The only way we can intervene on {Gibbard} is by intervening on Gibbard, and that is true at all philosophers’ contexts. So we should predict that across all these contexts an intervention on Gibbard will be more salient than one on {Gibbard}, and hence that insofar as we have given the matter any thought, we will commit to a conditional plan to intervene on Gibbard in order to intervene on {Gibbard}.

Furthermore, we should predict that across all these contexts we will *not* commit to intervening on {Gibbard} in order to intervene on Gibbard, since the latter is an intervention that we cannot perform, and which is therefore unlikely to be salient. So it is likely that at contexts like ours, individuals will typically not judge that Gibbard exists *because* {Gibbard} exists. This is, in fact, what we find when we look to philosophical judgements on this matter.

So our view can accommodate, and indeed explain, the kind of context-sensitivity that we see (and expect to see) regarding judgements of ⌜x *because* y⌝. In particular, we see converging judgements made by individuals who are likely to commit to the same interventions, and diverging judgements made by individuals who are likely to commit to different interventions.

That brings us to the third aspect of our practice in need of explanation. As we noted earlier, we sometimes (and philosophers often) judge that ⌜x *because* y⌝ even when no intervention on [y] is possible. If, as we contend, the core function of these judgments is to express commitment to/recommendation of, interventional norms, then such judgements appear mysterious. In what follows we argue that our tendency to make such judgements is a by-product of the sorts of capacities that we employ in order to make judgements in which we do express relevant non-cognitive attitudes.

Here is why that is. We have just said that in order to determine whether a judgement of ⌜x *because* y⌝ is apt we need to simulate which interventional norms we would be committed to at contexts other than our own. Where the context is one in which an intervention is possible, we are simulating *counterfactual* *interventions*. As we have noted, there is plenty of evidence that we are adept at such simulations.[[24]](#footnote-24) We think that we sometimes judge that ⌜x *because* y⌝ at contexts at which intervening on [y] is impossible because we (perhaps tacitly) run precisely the same kind of simulation in such contexts: we simulate a *counterpossible* intervention. The result of running such simulations tends to result in us making judgements of the relevant sort, even though those judgements do not express commitment to, or recommendation of, any interventional norm. All they do, in those contexts, is to assert a proposition that may be (minimally) true. But since sometimes all we care about are these minimal truths, we sometimes have reason to make such judgements.

The idea, then, is that when an intervention is impossible people sometimes nevertheless simulate to which interventional norms they *would* be committed, were they to occupy certain *counterpossible* contexts. In turn, they do this by simulating or imagining counterpossible interventions. They come to simulate these counterpossible interventions as a product of the very active mechanism that also allows them to simulate counterfactual interventions. In simulating these counterpossible interventions people are led to make the same sorts of judgements about whether ⌜x *because* y⌝, as they would make if they were instead simulating counterfactual interventions.

Two things stand in need of clarification at this point. First, we are not committed to the idea that as a matter of fact counterpossible conditionals (i.e., conditionals with counterpossible antecedents) take non-trivial truth-values. There are many philosophers who take this to be so (see Jago, 2015; Nolan, 2013 Baron, Colyvan & Ripley, 2017). But we require no such claim. We are merely committed to the idea that have a cognitive mechanism that allows us to simulate counterfactual interventions, and that mechanism does, in fact, also allow us to simulate counterpossible interventions and come to a view about what would happen were we to undertake some counterpossible intervention. That we are able to imagine counterpossible interventions and come to some view about what would happen, were we to perform one, is consistent with its in fact being the case that there are no non-trivially true counterpossible conditionals. There may be no fact of the matter what would happen if one were to perform a counterpossible intervention. All we require is that it seems to us as though there is such a fact.

It strikes us as immensely plausible that people have intuitions about what would happen were they to make certain counterpossible interventions. Philosophers clearly have such intuitions. They, after all, engage with counterpossible conditionals all the time, when they evaluate claims like ‘if 13 hadn’t been a prime number, then it would have been divisible by something other than itself and 1’ (Baker, 2005). Similarly, when we engage with claims like ‘if 13 hadn’t been a prime number, then cicadas wouldn’t have a 13-year life cycle’ we imagine an impossible world, in which 13 is not a prime number, and note that in that world cicadas have life cycles of 11 and 17.

We think the same is true of imagined counterpossible interventions. We do, as a matter of fact, tend to think that we can imagine counterpossible interventions. Indeed, participants in Latham & Miller’s (MSb) study appeared to have no difficulty in imagining intervening on what God approves of in order to intervene on what is good, and to imagine intervening on what is good in order to intervene on what God approves of. But even without such evidence the idea that we can simulate counterpossible inventions seem overwhelmingly plausible.

After all, it’s often not clear (even to philosophers) whether an intervention is possible or not. Is it impossible to intervene on whether water is H2O, or just very impractical? Certainly before Kripke came along plenty of philosophers thought it was possible. But the interventions they were imagining were, in fact, counterpossible ones, on the assumption that we now think that water is, of necessity, H2O. Quite generally, we, including non-philosophers, can be expected to often be unsure whether an intervention is possible or not. So it’s unsurprising that the cognitive mechanism that does the simulating is not very sensitive to whether the intervention is possible or not. It works just fine (at generating a simulation) regardless. Another way to put this point is to emphasise that the mechanism is effectively behind a hyperintensional veil, such that it does not know which interventions are possible and which not. It simply simulates interventions regardless of whether they are possible or not.

So, we say, the explanation of our tendency to make judgments of ⌜x *because* y⌝ even when there is no possible interventional norm to which we could be expressing commitment and recommendation, is that the mechanism that aids us in generating judgments about whether ⌜x *because* y⌝ in conditions in which we *do* express such commitments and recommendations often produces output (i.e., counterpossible simulations) which tend to lead us to make such judgements.

For this reason, we don’t think that our tendency to make judgments of ⌜x *because* y⌝ even when there is no possible interventional norm to which we could be expressing commitment and recommendation undermines the plausibility of non-cognitivism. In our view, the core function of the discourse in which such judgements are embedded is to express commitment to, and recommendation of interventional norms at contexts. The fact that the mechanism we use to make these judgements also leads us to make judgements of this kind in contexts where the relevant intervention is impossible—and thus cannot plausibly be recommended—is entirely compatible with the discourse having this core function. It is simply that these particular judgements do not fulfil the core function of judgements of their kind, or contribute to the core function of the discourse in which they are embedded. It is in this sense that judgements that do not express these non-cognitive attitudes but only assert minimally truth-apt propositions are parasitic on judgements that do. This is not to say that there is anything wrong with making these assertions, or that they are sub-par in some way. Indeed, sometimes we are interested in evaluating such assertions, especially in philosophy. Our claim is simply that expressing truth-apt propositions is not the core function of judgements of ⌜x *because* y⌝.

**5. Conclusion**

This paper argues that non-cognitivism and non-descriptivism about metaphysical explanation is a genuine rival to cognitivism. It makes good sense of our practices surrounding metaphysical explanation. None of this is to say that the view does not face obstacles that we have not considered here; but we hope to have said enough to motivate further consideration of this view.

**References**

Ashley, J., & Tomasello, M. (1998). ‘Cooperative problem‐solving and teaching in preschoolers.’ *Social Development*, *7*(2):143-163.

Audi, P. (2012). ‘A clarification and defense of the notion of grounding.’ In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding: Understanding the structure of reality* (pp. 101–121). Cambridge: CUPress.

Baker, A., 2005, “Are There Genuine Mathematical Explanations of Physical Phenomena?”, Mind, 114(454): 223–238.

Baron, S., Colyvan, M., & Ripley, D., (2017). ‘How Mathematics Can Make a Difference’, *Philosophers’ Imprint*, 17(3):1–29.

Baron, S., & Norton, J.(2019). ‘Metaphysical Explanation: The Kitcher Picture.’ *Erkenntnis*. DOI 10.1007/s10670-018-00101-2

Chin-Parker, S., & Bradner, A. (2010). ‘Background shifts affect explanatory style: How a pragmatic theory of explanation accounts for background effects in the generation of explanations.’ *Cognitive processing*, *11*(3):227-249.

Daly, C. (2012). ‘Skepticism about Grounding.’ In F. Correia and B. Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: CUP.

Dasgupta, S. (2017). ‘Constitutive Explanation.’ *Philosophical Issues* 27(1):74-97.

Duncan, M., Miller, K & Norton, J. (2017). ‘Is Grounding a Hyperintensional Phenomenon?’ *Analytic Philosophy,* 58(4), 297-329.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*(1).

Flocke, V. (2020). Carnap’s Noncognitivism about Ontology. *Noûs* 54 (3):527-548.

Flocke, V. (forthcoming). ‘Ontological Expressivism.’ In J. Miller (ed.) *The Language of Ontology*. Oxford University Press.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgement.* Harvard University Press.

Gibbard, A. (2003). *Thinking How to Live.* Harvard University Press.

Glazier, M. (2020). Explanation. In Michael J. Raven (ed.), *Routledge Handbook of Metaphysical Grounding*. London: pp. 121-132.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., and Danks, D. (2004). ‘A theory of causal learning in children: Causal maps and Bayes nets.’ *Psychological Review,* 111:1–30.

Hilton, D. J. (1990). ‘Conversational processes and causal explanation.’ *Psychological Bulletin*, *107*(1):65.

Hilton, D. J., & Erb, H. P. (1996). ‘Mental models and causal explanation: Judgements of probable cause and explanatory relevance.’ *Thinking & Reasoning*, 2(4):273–308.

Hilton, D. J., & Slugoski, B. R. (1986). ‘Knowledge-based causal attribution: The abnormal conditions focus model.’ *Psychological review*, 93(1), 75.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy*, *106*(11):587-612.

Jago, M. (2015). ‘Impossible worlds.’ *Noûs*, *49*(4):713-728.

Jansson, L. (2018). ‘When are structural equation models apt? Causation versus grounding’ In Reutlinger, A., & Saatsi, J. (Eds.) *Explanation beyond causation: Philosophical perspectives on non-causal explanations*. Oxford University Press.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review*, *93*(2):136.

Kelemen, D., Rottman, J., & Seston, R. (2013). ‘Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default.’ *Journal of Experimental Psychology: General*, 142(4):1074.

Knobe, J. (2009). Folk judgments of causation. *Studies in History and Philosophy of Science Part A*, *40*(2), 238-242.

Kovacs, D. M. (2017). ‘Grounding and the argument from explanatoriness.’ *Philosophical Studies*, *174*(12):2927-2952.

Koslicki, K. (2015). ‘The Coarse-Grainedness of Grounding.’ in K. Bennett and D. Zimmerman (ed.), *Oxford Studies in Metaphysics*, Oxford: Oxford University Press

Kushnir, T., Gopnik, A., Lucas, C. & Schulz, L. (2010). ‘Inferring Hidden Causal Structure.’ *Cognitive Science,* 34(1):148-160

Lagnado, D. A., and Sloman, S. A. (2004). ‘The advantage of timely intervention.’ *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 30:856–876.

Latham, A. J. & Miller, K. (MSa) ‘Judgements of Metaphysical Explanations are Context Sensitive.’ <https://philarchive.org/rec/LATJOM>

Latham, A. J. & Miller, K. (MSb) ‘Accommodating Our Pattern of Judgements about Metaphysical Explanations: Evidence that Metaphysical Explanation is not Asymmetric’ <https://philpapers.org/rec/LATAOP>

Lombrozo, T. (2007). ‘Simplicity and probability in causal explanation.’ *Cognitive psychology*, *55*(3):232-257.

Maurin, A-S. (2018). ‘Grounding and metaphysical explanation: it’s complicated.’ *Philosophical Studies*. DOI:10.1007/s11098-018-1080-0

McGill, A. L. (1989). ‘Context effects in judgments of causation.’ *Journal of Personality and Social Psychology*, *57*(2):189.

Menzies, P. (2007). Causation in context. *Causation, physics, and the constitution of reality*, 191-223.

Miller, K & Norton, J. (2017). ‘Grounding: it’s (probably) all in the head.’ *Philosophical Studies*, 174(12):3059-3081.

Murphy, G. L., & Medin, D. L. (1985). ‘The role of theories in conceptual coherence.’ *Psychological review*, *92*(3):289.

Nolan, D. P. (2013). Impossible worlds. *Philosophy Compass*, *8*(4), 360-372.

Norton, J., & Miller, K. (2019). ‘A Psychologistic Theory of Metaphysical Explanation’. *Synthese*, *196*(7):2777-2802.

Pearl, Judea (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, *49*(1-2):123-163.

Price, H. (2001). ‘Causation in the special sciences: the case for pragmatism.’ In M. C. Galavotti, P. Suppes and D. Constantini. (eds.) *Stochastic causality.* CSLI Publications. pp. 103-120.

Rhodes, M., Bonawitz, E., Shafto, P., Chen, A., & Caglar, L. (2015). ‘Controlling the message: preschoolers’ use of information to teach and deceive others.’ *Frontiers in psychology*, *6*.

Rodriguez-Pereyra, G. (2005). ‘Why Truthmakers?’ In H. Beebee and J. Dodd (eds.), *Truthmakers: The Contemporary Debate*. Oxford: Clarendon Press.

Rosen, G. (2010). ‘Metaphysical Dependence: Grounding and Reduction’. In *Modality: Metaphysics, Logic, and Epistemology*, edited by B. Hale and A. Hoffmann, pp. 109–36.

Schaffer, J. (2009). ‘On What Grounds What.’ In D. Manley, D. Chalmers, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: OUP.

Schaffer, J. (2010). ‘The Internal Relatedness of All Things’ *Mind* 119:341-376.

Schaffer, J. (2016). ‘Grounding in the image of causation.’ *Philosophical Studies* 173 (1):49-100.

Shaheen, J. L. (2017). ‘The causal metaphor account of metaphysical explanation.’ *Philosophical Studies*, *174*:553–578.

Sloman, S. (2005). *Causal Models: how people think about the world and its alternatives*. Oxford: OUP.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). ‘Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers.’ *Cognitive Science*, 28:303–333.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive science*, *27*(3), 453-489.

Thompson, N. (2016). Grounding and metaphysical explanation. In *Proceedings of the Aristotelian Society* (Vol. 116, No. 3, pp. 395-402). Oxford University Press.

Thompson, N. (2019). Questions and Answers: Metaphysical Explanation and the Structure of Reality. *Journal of the American Philosophical Association*, *5*(1), 98-116.

Tomasello, M., Davis-Dasilva, M., Cama, K, L., & Bard, K. (1987). ‘Observational learning of tool-use by young chimpanzees.’ *Human evolution*, *2*(2):175-183.

Trogdon, K. (2013). An introduction to grounding. In Miguel Hoeltje, Benjamin Schnieder & Alex Steinberg (eds.), Varieties of Dependence. Munich, Germany: Philosophia Verlag. pp. 97-122.

Van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: OUP.

Vasilyeva, N., Wilkenfeld, D., & Lombrozo, T. (2017). ‘Contextual utility affects the perceived quality of explanations.’ *Psychonomic bulletin & review*, *24*(5):1436-1450.

Vlach, H. A., & Noll, N. (2016). Talking to children about science is harder than we think: characteristics and metacognitive judgments of explanations provided to children and adults. *Metacognition and Learning*, *11*(3):317-338.

Wilson, A. (2018a). ‘Metaphysical causation.’ *Noûs*, *52*(4):723-751.

Wilson, A. (2018b). ‘Grounding Entails Counterpossible Non‐Triviality.’ *Philosophy and phenomenological research*, *96*(3), 716-728.

Wilson, J. (2014). ‘No Work for a Theory of Grounding.’ *Inquiry*, 57(5–6):1–45.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: OUP.

1. Corner quotes signify that ⌜x *because* y⌝ is a *kind* of sentence, where x and y are variables that range over sentences. We will speak of ‘an instance of ⌜x *because* y⌝’ when we intend to talk about a particular instance of the schema. We will simply speak of ⌜x *because* y⌝ in order to talk about all instances of the schema. [↑](#footnote-ref-1)
2. Though Wilson (2018a; 2018b) and Schaffer (2016) think metaphysical explanation is a sort of non-diachronic causal explanation. [↑](#footnote-ref-2)
3. [x] is read as ‘the fact that x’. Facts are structured parts of the world comprised of objects, properties and relations. [↑](#footnote-ref-3)
4. Schaffer (2009), Audi (2012), and Rodriguez-Pereyra (2005) endorse truth-conditions along these lines. [↑](#footnote-ref-4)
5. Exceptions include Dasgupta (2017), Norton & Miller (2019), Wilson (2014), Shaheen (2017), and Baron & Norton (2019). [↑](#footnote-ref-5)
6. Radical, but not unheard of: Flocke (forthcoming) motivates and proves the coherence of *ontological expressivism* as a productive way of understanding debates in ontology, and Flocke (2020) argues that a class of Carnapian external statements (*pragmatic* external statements) express non-cognitive dispositions. [↑](#footnote-ref-6)
7. We take contexts to include an individual at a particular spatiotemporal location in a world. Hence contexts include psychological and epistemic facts about the individual, including their goals. [↑](#footnote-ref-7)
8. Gopnik et al., (2004), Steyvers, Tenenbaum, Wagenmakers & Blum (2003) and Sloman (2005). [↑](#footnote-ref-8)
9. Einhorn & Hogarth (1986) and Murphy & Medin (1985). [↑](#footnote-ref-9)
10. Lagnado & Sloman (2004). [↑](#footnote-ref-10)
11. Interventionist accounts of causation have been provided by Pearl (2000), Woodward (2003) and Menzies (2007). [↑](#footnote-ref-11)
12. Kushnir, Gopnik, Lucas & Schultz (2010), Lagnado & Sloman (2004). [↑](#footnote-ref-12)
13. Sloman (2005). [↑](#footnote-ref-13)
14. Though Jansson (2018) points to difficulties determining which such models are apt, in the case of grounding. [↑](#footnote-ref-14)
15. Of course, Gibbard’s (2003) eventual view is somewhat more complicated than this. He ultimately frames his view of the content of the relevant mental states in terms of sets of world/hyperplan pairs, where *hyperplans* are maximally detailed contingency plans regarding what to do in any possible circumstance. For our purposes the simpler notion of a conditional plan will suffice. [↑](#footnote-ref-15)
16. Rhodes, Bonawitz, Shafto, Chen & Caglar (2015). [↑](#footnote-ref-16)
17. Sobel, Tenenbaum & Gopnik (2004) and Ashley & Tomasello (1998). [↑](#footnote-ref-17)
18. For example, see Tomasello, Davis-Dasilva, Cama & Bard (1987). [↑](#footnote-ref-18)
19. Indeed, these cases of overdetermination are just one species of this kind of case. We can imagine cases in which there is some ‘blocker’ such that although one fact depends on another, there is, in a particular world, no intervention that one can in fact perform that will succeed in intervening on the dependent fact, since any such intervention will be ‘blocked’ by the presence of the blocker. Of course, it’s still true that had the blocker not been present, the intervention would have succeeded. So we can still model an intervention here by counterfactually holding fixed the absence of blocker. But the point is that nothing we can in fact do, in the world in question, will result in us successfully intervening on one fact via the other, since in the world in question the blocker cannot (we are supposing) be removed. (If it helps, you can think of the blocker as an all-powerful demon). [↑](#footnote-ref-19)
20. With thanks to an anonymous referee for these particular examples. [↑](#footnote-ref-20)
21. It is also important to remember that much of recent theorising about metaphysical explanation is not directed at making, or evaluating, particular judgements of this form. Often it involves theorising about what makes such judgements true (on the assumption that some are) and, in turn, it often involves reporting such judgements. Unsurprisingly, when philosophers report judgements of ⌜x *because* y⌝ they are usually not best seen as committing to particular interventional norms, nor recommending any such norms. Importantly then, our account is not intended to be an account of what these philosophers are doing. In what follows we focus only on the first-order judgements that philosophers make. [↑](#footnote-ref-21)
22. Hilton & Slugoski (1986), Knobe (2009), Hitchcock & Knobe (2009) and Kahneman & Miller (1986). [↑](#footnote-ref-22)
23. Hitchcock & Knobe (2009) and Kahneman & Miller (1986). [↑](#footnote-ref-23)
24. See the references in fn. 12 and fn. 13. [↑](#footnote-ref-24)