

20

Mental and normative causation

Teemu Toppinen & Vilma Venesmaa

Like cats, we often do what we do simply because we want to. Sometimes we might even choose a certain course of action because we think that it is the right thing to do. Our joy tends to be contagious and may evoke smiles and laughter; on occasion we may wake up, in the middle of the night, due to a terrifying nightmare. The impact of the mental on the physical is a pervasive and a very familiar phenomenon. Yet it is also, of course, a very familiar fact that this phenomenon gives rise to a number of philosophical puzzles, some of which we often lump together under the heading of *mental causation*. In his work on the topic, Panu Raatikainen (2006, 2007, 2010, 2013, 2018) offers an account of mental causation, drawing on an interventionist approach to causation – developed, especially, in the context of philosophy of science – and on the idea that causal claims would carry an (often) implicit reference to contrast classes.

Raatikainen notes that while he focuses on the case of mental causation, his conclusions “are applicable across the board” in relation to the special sciences. This, it seems, would be a very nice feature of the account. However, we argue that if the kind of account that Raatikainen proposes is correct, then this has implications not just for the special sciences, but also for the case of the *normative*. The Raatikainen-style account makes it relatively easy for normative properties such as rightness, wrongness, being supported by the balance of reasons, having aesthetic merit, or being morally depraved, to have causal power. In particular, we suggest that the following conditional is true:

CONDITIONAL. If the interventionist account of mental causation of the kind that Raatikainen proposes is correct, then normative properties have causal power, even given a non-naturalist or a quasi-realist understanding of such properties.

We suggest, moreover, that normative properties should not turn out to have causal power given a non-naturalist or a quasi-realist construal of such properties. And so, the truth of CONDITIONAL turns out to be problematic for the style of account of causation that Raatikainen favors. We start by presenting Raatikainen's proposal with regard to how to make sense of mental causation (§§1–3). We then explain why the account generalizes to the case of the normative properties (§§4–5) and briefly conclude (§6).

§1. Following Raatikainen (2010, sec. 2; Raatikainen, on his part, follows Bennett 2007 here), we may approach the problem of mental causation in the context of what is known as *the exclusion problem*. Very briefly, the problem is that the following claims all seem very plausible, yet incompatible:

- (1) DISTINCTNESS. Mental properties (and perhaps events) are distinct from physical properties (or events) (i.e., the type-identity theory is false).
- (2) COMPLETENESS. Every physical occurrence has a sufficient physical cause.
- (3) EFFICACY. Mental events sometimes cause physical events and sometimes do so in virtue of their mental properties.
- (4) NO OVERDETERMINATION. The effects of mental causes are not systematically overdetermined.
- (5) EXCLUSION. No effect has (at a particular time *t*) more than one sufficient cause unless it is overdetermined.

Different solutions have been proposed to the problem (see Bennett 2007), but our focus is solely on the one offered by Raatikainen (2006, 2007, 2010, 2013, 2018).¹ Raatikainen's solution takes issue with the way the entire problem is set up. On his account, the exclusion problem is based on a confusion, really, or on a failure to adequately understand the notion of a cause. As a result, Raatikainen rejects – or at least refuses to accept – all of (2), (4), and (5), as these theses are formulated above. *Prima facie*, this of course doesn't look very promising. As noted, the whole problem arises because all of (1)–(5) are all very attractive theses. But Raatikainen's idea is to explain why we can let go of the relevant theses, once we properly understand the concept of causation. We next turn to Raatikainen's account of this concept.

¹ The references to Raatikainen's work, below, are all to the 2010 paper.

§2. Raatikainen's proposal is an application of an *interventionist* approach to causation.² The fundamental – and quite sensible – thought behind the interventionist approach is to start by asking what the point of our having the concept of causation is in the first place. What role or function does this concept perform? What are we up to when we make judgments about what causes what? The answer, very roughly, is that

[...] knowledge of genuine causal relationships is, *sometimes*, practical and applicable: by manipulating the cause we can influence the effect. If there is a real causal relationship between *A* and *B*, manipulating *A* is a way to change *B*; Mere correlation between *C* and *D*, on the other hand, just disappears if one attempts to affect *D* by manipulating *C* (p. 353).

The interventionist view “connects causal claims with counterfactual claims concerning what would happen to an effect under interventions on its putative cause” (p. 353):

Roughly, *C* causes *E* if and only if an intervention on *C* would bring about a change in *E*. Slightly more exactly, causal claims relate, in this approach, variables, say *X* and *Y*, that can take at least two values. These may often be some magnitudes (such as temperature, electric charge or pressure), but in simple cases, they may also be just discrete alternative events or states of affairs. The idea now is that were there an intervention on the value of *X*, this would also result [in] a change in the value of *Y* (p. 353).

The interventionist view, then, provides truth-conditions of roughly the following kind for claims about what causes what:

A change in *X* causes a change in *Y* if and only if, if *X* were to be changed by an intervention to such and such a value, the value of *Y* would change.

Raatikainen subscribes to the popular idea that “causal claims do not in fact describe a simple binary relation between two events, but rather involve (even if often only implicitly) a contrastive class for both cause and effect” (p. 354). The contrast classes are contextually determined, but given “default contrast classes,” where the alternatives for *X* or *Y* having a certain value simply is their not having that value, (see pp. 354–355), we then get:

X's being *x_i*, (rather than not being *x_i*) causes *Y*'s being *y_i*, (rather than not being *y_i*) if and only if, if *X*'s being *x_i* were to be changed by an intervention to *X*'s not being *x_i*, then *Y* would change from being *y_i* to not being *y_i*.

² Raatikainen provides helpful references to the literature in which the interventionist theory has been developed, giving special credit to the work of James Woodward (see, e.g., Woodward 2003).

§3. According to Raatikainen, the interventionist account implies that a mental state can be a cause of behavior, and, given certain assumptions about the contrast classes, the physical state that underlies the relevant mental state may fail to be such a cause. Raatikainen (pp. 355–356) argues for this by appealing to the case of John who desperately wants a beer.³ Happily, John remembers having earlier bought a six-pack and having placed it in the refrigerator. He then forms a belief that there is some beer in the refrigerator and consequently walks to the refrigerator. What causes John's walking to the refrigerator? Is it his belief that there is some beer in the fridge, or perhaps his brain state, *B*, at the moment – the one underlying his belief?

Given the default contrasts, it turns out that John's belief causes his behavior, but his brain state doesn't. Consider, first:

- (i) If John's belief that there is beer in the refrigerator were to be changed by an intervention to not having the belief, he would not have gone to the refrigerator.

According to the standard possible world analysis of counterfactual conditionals, ' $P \rightarrow Q$ ' is true if and only if either there is no *P*-world, or some *P* & *Q* -world is more similar to the actual world than any *P* & not-*Q* -world. In the case of (i), it is not the case that there are no *P*-worlds. But there are *P* & *Q* -worlds that are closer than any *P* & not-*Q* -worlds. Intuitively, the idea is that if we just change John's belief state, so that he does not believe that there is beer in the refrigerator, then he does not go to the refrigerator (but, say, goes to the grocery instead). If we would want to make it true that John's belief state changes, but that John nevertheless goes to the refrigerator (and not to the grocery, for example), then more changes would be needed to the way things actually are, and we thus move to possible scenarios that are further from the actual world. (i), then, plausibly comes out as true. Consider next:

- (ii) If John's brain state *B* were to be changed by an intervention to his not having that state, he would not have gone to the refrigerator.

In the case of (ii), it is not true that all of the *P* and not-*Q* -worlds are further away from the actual world than some *P* and *Q* -world. Given that belief states are multiply realizable by different brain states, there are possible worlds in which John's brain state *B* is changed to a different brain state, *B'*, but in which John nevertheless goes to the refrigerator. Also, these worlds would seem to be closer to the actual world than the worlds in which John's brain state is manipulated in the relevant way, but he goes to the grocery, say, instead of heading to the refrigerator. And so, (ii) plausibly is not true.

According to Raatikainen's interventionist proposal, again:

³ Raatikainen notes that his argument has been inspired by Tim Crane's (2001) "similar argument with respect to a more traditional counterfactual approach to causation" (p. 358, n. 15).

X 's being x_i (rather than not being x_i) causes Y 's being y_i (rather than not being y_i) if and only if, if X 's being x_i were to be changed by an intervention to X 's not being x_i , then Y would change from being y_i to not being y_i .

In the light of the above, if we now replace X 's being x_i with John's believing that there is beer in the refrigerator, X 's not being x_i with John's not believing that there is beer in the refrigerator, Y 's being y_i with John's going to the refrigerator, and Y 's not being y_i with John's not going to the refrigerator (but going to the grocery instead), we get the result that John's belief causes his going to the refrigerator. By the interventionist analysis, the causal power of John's belief is vindicated.

By contrast, if we replace X 's being x_i with John's being in a certain brain state and X 's not being x_i with John's not being in this brain state (but in some other brain state instead), we get the perhaps somewhat surprising result that John's brain state does not cause his action. This is not to say, Raatikainen emphasizes, that John's being in the relevant state is not a *sufficient condition* for his performing the action in question. However, by Raatikainen's analysis, it is not a *cause* of his action: "Being sufficient condition for the occurrence of something, and being its difference-making cause, must thus be clearly distinguished" (p. 358).

Raatikainen notes that the argument that he gives "certainly deserves, and requires, further elaboration" (p. 358). For instance, in order for John's belief to cause his action, it must be the case that the alleged potential interventions are *genuine interventions*. They must not directly cause the change in John's behavior, for example. So, the mere truth of conditionals such as (i) and (ii) does not, strictly speaking, suffice for establishing the causal power of John's belief. The account is, in fact, more complicated. But Raatikainen suggests that the extra complexities will not cause any trouble for his argument (for a brief discussion of this, see pp. 358–359). We see no reason to suspect that this would not be so, and grant this assumption here. That is, we grant that by the standards of the interventionist theory, John's belief gets to cause his going to the refrigerator.

What is essential, for our purposes here, is that Raatikainen's account vindicates the causal efficacy of the mental. But as Raatikainen's account has been presented in the context of the exclusion problem, we may briefly note what Raatikainen takes to be the implications of his proposal in relation to the problem. Let us consider, then, in the light of Raatikainen's view, the exclusion problem again. In particular, consider (2) and (5):

- (2) COMPLETENESS. Every physical occurrence has a sufficient physical cause.
- (5) EXCLUSION. No effect has (at a particular time t) more than one sufficient cause unless it is overdetermined.

According to Raatikainen (p. 360), both of these assumptions involve confusing causes with sufficient conditions:

There are causes, which are difference-makers; and there are sufficient conditions, which are wholly different issues and not causes of any sort; there are no such things as *sufficient causes*. Hence, I do not think that these two assumptions are so much false (or true) as mongrels based on a conceptual confusion which fail to make clear sense.

If we try to reformulate these theses in terms of difference-making causes, the resulting theses turn out to be false. The revised version of (2), for instance, would state that every physical occurrence has a physical difference-making cause, but Raatikainen suggests that his example of John establishes that this is mistaken (p. 361). (Raatikainen (p. 360) suggests that (4), or the thesis that the effects of mental causes are not systematically overdetermined, also involves confusion, but we shall not delve into this issue here).

However, again, what is really essential, for our purposes, is the way in which Raatikainen's account allows for the possibility of the mental getting some real causal work done. We next turn to metaethics and address the way in which Raatikainen's ideas generalize to the realm of the normative.

§4. Moral properties plausibly play a role in causal *explanations*. A helpful pair of examples may be lifted from a classic paper by Nicholas Sturgeon. First, we'll play the Nazi card: it is plausible that Hitler initiated a world war and ordered the "final solution" at least in part because he was morally depraved (Sturgeon 1984, p. 249). Second:

An interesting historical question is why vigorous and reasonably widespread moral opposition to slavery arose for the first time in the eighteenth and nineteenth centuries, even though slavery was a very old institution; and why this opposition arose primarily in Britain, France, and in French- and English-speaking North America, even though slavery existed throughout the New World. There is a standard answer to this question. It is that chattel slavery in British and French America, and then in the United States, was much *worse* than previous forms of slavery, and much worse than slavery in Latin America (Sturgeon 1984, p. 245).

The second explanation, at least, is undoubtedly controversial. But this is irrelevant. What these examples are meant to illustrate is simply the fact that in giving causal explanations of various events, it is sometimes quite natural to appeal to certain things having certain normative (e.g., moral) properties. This would seem to be so because it is plausible that counterfactuals such as the following are true:

HITLER. Had Hitler not been morally depraved, he would not have initiated a world war and ordered the "final solution."

SLAVERY. Had the slavery in British and French America, and then in the United States, not been worse than previous forms of slavery, and slavery in Latin America, vigorous and reasonably widespread moral opposition to slavery would not have arisen in the way it did.

Just consider the possibility that Hitler would not have been morally depraved, but would rather have been “humane and fair-minded, free of nationalistic pride and racial hatred” (Sturgeon 1988, p. 250). In this case, he would not have done what he did. Or consider the possibility that he would have been, not morally depraved, but just a bit of a jerk. In this case, too, he plausibly would not have done what he did. Similar considerations apply to SLAVERY.

We may also give these counterfactuals an interventionist twist:

HITLER*. Had Hitler’s character been changed by an intervention from being morally depraved to not being morally depraved, he would not have initiated a world war and ordered the “final solution.”

SLAVERY*. Had the slavery in British and French America, and then in the United States, been changed by an intervention to not having been worse than previous forms of slavery, and slavery in Latin America, vigorous and reasonably widespread moral opposition to slavery would not have arisen in the way it did.

HITLER* and SLAVERY* also seem to be true. The mere addition of the relevant changes being brought about by an intervention makes no difference.

We make the simplifying assumption, here, that given the interventionist view, establishing that normative properties support counterfactuals such as HITLER, HITLER*, SLAVERY, and SLAVERY* suffices to establish the causal efficaciousness of the normative. We were willing to grant, above, that whatever further conditions must apply (e.g., regarding the alleged interventions really being genuine interventions), do apply in the case of the mental. We now assume that this plausibly is the case also in the case of the normative. That is, we see no reason to think that there would be any significant difference between the cases of mental and normative causation, with respect to the complications of the relevant sort. Given the assumption that the cases are analogous, in this respect, the truth of HITLER* and SLAVERY* establishes, by the lights of Raatikainen’s account of causation, that Hitler’s moral depravity and the level of badness of the slavery in British and French America, and then in the United States, are properties with causal power.

§5. According to Raatikainen’s view, then, it is quite easy for normative properties to be causally efficacious. They just need to support the truth of counterfactuals such as HITLER, HITLER*, SLAVERY, and SLAVERY* (and pass whatever further conditions the interventionist theory involves, which we are now supposing they do pass).

Interestingly, it seems that their supporting such counterfactuals can be made sense of on a wide variety of different views in metaethics.

Consider, first, *non-naturalist* views. These views are *representationalist* in that according to these views, normative judgments represent, in some substantive and theoretically interesting sense, the ways the world is or, as we could also put it, the ways in which normative properties are instantiated. Moreover, according to non-naturalism, these properties are *sui generis*, irreducibly normative properties. On this view, normative properties such as the properties of being morally depraved or being very bad are something very different from the *natural* properties, which are properties such that can figure in empirical regularities and are amenable to study by empirical science. (For examples of non-naturalist views, see, e.g., Enoch 2011; Bengson, Cuneo & Shafer-Landau 2024.)

However, even though normative properties are, according to the non-naturalist, very different from the natural properties, the normative doesn't, even on the non-naturalist account, float free of the natural – to deploy a much-used phrase from Simon Blackburn (1984, p. 221). Rather, despite normative properties being so very different from natural properties, the normative ways the world is nevertheless *supervene* on the natural ways the world is, in something like the following way:

SUPER. Metaphysically necessarily, for all x and all properties F in the set of normative properties N , if x has F , then there is some G in the set of natural properties D such that x has G , and metaphysically necessarily, for all y , if y has G , it has F .

It is metaphysically impossible, then, that there would be a normative difference between two possible things without there being some difference in their natural properties; it is metaphysically necessary that the natural ways the world is fix – with metaphysical necessity – the normative ways the world is.⁴

A similar assumption – implicit, in the discussion, above – about the supervenience of the mental on the physical, is crucial to the interventionist account of mental causation. In motivating the problem of mental causation, it was assumed that mental properties are not simply identical with physical ones. But, in order for it to be possible to manipulate the physical via interventions on the mental, it must nevertheless be assumed that the mental supervenes on the physical.

Now, it is an excellent question how non-naturalists can explain the supervenience of the normative on the natural (for an overview of the issues that this raises for the non-naturalist, see Väyrynen 2018). But what is relevant here is that non-naturalists, in any case, *accept* that the normative supervenes on the natural. We then have a guarantee, given this kind of non-naturalist view, that had Hitler not been morally

⁴ Some non-naturalists reject SUPER because they deny that it is metaphysically necessary that the natural ways the world is fix, with *metaphysical necessity*, the normative ways the world is. These non-naturalists believe that the natural ways the world is only determine the normative ways the world is with a weaker *normative necessity*. For this kind of view, see, e.g., Rosen 2020. We set these views aside here.

depraved, his natural properties, too, would have been different. Given a suitable story about the kinds of natural properties that moral depravity supervenes on, we then have a guarantee that had Hitler not been morally depraved, he would have been free of the kind of combination of nationalistic pride and racial hatred that he sadly manifested. This allows normative properties to support, even given a non-naturalist understanding of such properties, the sorts of counterfactuals that would make it true, on the interventionist account of causation, that Hitler's moral depravity caused his initiating a world war and ordering the "final solution."

It seems, then, that according to the kind of account of causation favored by Raatikainen, the irreducibly normative *sui generis* properties postulated by the non-naturalist view have causal power. This seems like an interesting consequence for a theory of causation to have. For standardly, perhaps, non-natural properties are understood to be properties such that do *not* have causal power (see, e.g., Enoch 2011). Indeed, not having causal power would be one candidate for a feature that makes these properties non-natural. (For an account that grants that non-natural normative properties have causal powers, see Oddie 2005; for a recent attempt at navigating the complexities involved in giving a good account of what it is for a property to be non-natural, see Leary 2021).

Another type of view that we wish to highlight here is *quasi-realism*. Quasi-realism is a *non-representationalist, expressivist*, view, according to which normative language does not represent, in any substantive or theoretically interesting sense, normative properties and facts. Instead, on the expressivist view, normative language has a "dynamic," broadly practical, meaning. Its function is to guide attitude-formation and action. Very roughly, according to expressivist views, judging that Hitler was morally depraved is not a matter of representing Hitler as having had a certain kind of specifically normative property. Or at least this is not a very illuminating way of understanding the nature of this kind of judgment. Instead, judging that Hitler was morally depraved is more helpfully, and more fundamentally, understood in terms of being committed to acting or feeling in certain ways – in terms of being committed to disapproving of Hitler's character, say. (For this kind of way of understanding expressivism, see Dreier 2004, 2015; for expressivist views, see, e.g., Blackburn 1998, Gibbard 2003, Ridge 2014.)

Quasi-realism is what we get when we combine an expressivist view about the meaning of normative language and the nature of normative thought with the idea that there are normative properties, truths, and facts. According to the quasi-realist, some actions really have the property of being right, while others have the property of being wrong; it really is true – a fact – that Hitler was morally depraved; and so on. (Moreover, according to quasi-realism, such properties, truths, and facts are objective, in a certain interesting sense, but we can set this issue to one side here).

Now, it is an interesting question how expressivists can earn the right – to use another familiar phrase of Blackburn's – to accept the existence of normative properties, truths, and facts. A part of the reply is often taken to be that properties, truths, and facts don't carry a steep metaphysical price. For Hitler to have the

property of being morally depraved, and for it to be a truth, or a fact, that Hitler was morally depraved, just is for it to be the case that Hitler was morally depraved. That Hitler was morally depraved is a normative claim that seems entirely compatible with quasi-realism. And so, it also seems entirely compatible with quasi-realism that there are normative properties (e.g., moral depravity, which Hitler instantiated), truths, and facts (e.g., the truth and the fact that Hitler was morally depraved). In any case, what is relevant here is just that quasi-realists accept the existence of normative properties, truths, and facts. (Almost all expressivists are quasi-realists, these days; see, again, e.g., Blackburn 1998, Gibbard 2003, Ridge 2014.)

Quasi-realists also accept that the normative properties of things supervene on their natural properties. Again, there are interesting questions that may be asked about whether they can really make good sense of the truth of something like *SUPER*. (They can.) But what is relevant here is that quasi-realists, in any case, accept the supervenience claim.

Given that quasi-realists accept the existence of normative properties, and that these properties supervene on the natural, quasi-realists, too, are in a position to accept claims such as *HITLER** and *SLAVERY**. Consequently, it seems that on the assumption of the Raatikainen-style account of causation, quasi-realists, too, get to have normative properties with causal power. As was the case with non-naturalism, here, too, the standard view presumably is that normative properties, as understood by the quasi-realist, are *not* the kind of thing that would have causal power. Quasi-realists have tried to explain how they, too, can make sense of *causal explanations* with normative explanantia (Blackburn 1991, Gibbard 2003, Sinclair 2012). But the idea has been to just explain the causal *relevance* of normative properties, or the relevance of normative properties to causal explanation. One possibility, here, is to appeal to the idea that when something has a normative property, this non-causally secures its also having some natural properties, which, in turn, can then do the *causing* (for this kind of account of macro-level causal explanation, more generally, see Jackson & Pettit 1990). But it is often thought that for a quasi-realist, normative properties are merely ‘shadows of predicates’ – that for a quasi-realist, being morally depraved just amounts to being a thing such that the predicate ‘morally depraved’ may truthfully be applied to it. And it would be natural to think that things are not imbued with causal power in virtue of having such shadowy properties, even if their having these shadow properties may be helpfully appealed to in giving causal explanations.⁵

⁵ It is not clear that quasi-realists should treat normative properties as mere shadows of predicates. Some quasi-realists accept a reductive view, according to which normative properties are natural properties. It’s just that claims about which natural properties normative properties reduce to are given an expressivist interpretation (for this kind of view, see, e.g., Bex-Priestley 2024). Given a reductive quasi-realist account, the idea of a normative property as being causally efficacious does not sound so puzzling. But even on this kind of expressivist view, some claims about what causes what, namely the claims about normative causation, turn out to have a non-representational meaning, which some might find surprising. In any case, the Raatikainen-style account of causation seems to entail that normative properties need not be causally inert even if they are mere shadows of normative predicates that play a non-representational role.

§6. CONDITIONAL says, again, the following:

CONDITIONAL. If the interventionist account of mental causation of the kind that Raatikainen proposes is correct, then normative properties have causal power, even given a non-naturalist or a quasi-realist understanding of such properties.

We have suggested that CONDITIONAL is plausibly true. We have also suggested, in the previous section, that normative properties do *not* have causal power, given a non-naturalist or a certain kind of quasi-realist understanding of such properties. If this is true, then this entails, together with CONDITIONAL, that the kind of interventionist account of causation defended by Raatikainen is not correct.

Is there a way for a defender of the interventionist account to resist this line of argument? One way to do so would be to turn our *modus tollens* into a *modus ponens* and suggest that given the plausibility of the interventionist account of causation, non-natural normative properties or normative properties as shadows of predicates should be construed as causally efficacious. This strikes us as an unpromising strategy. Another possible line of response would be to reject our assumption that there are no significant asymmetries between the cases of the mental and the normative, when it comes to the applicability of the interventionist account (e.g., the conditions that something must satisfy in order to be a genuine intervention). While we are not optimistic about the prospects of finding asymmetries of a relevant kind, a defense of the assumption that none exist remains outside the scope of this paper.

References

- Bengson, John, Cuneo, Terence & Shafer-Landau, Russ (2024): *The Moral Universe*. Oxford: Oxford University Press.
- Bennett, Karen (2007): 'Mental Causation', *Philosophy Compass* 2: 316–337. URL = <https://doi.org/10.1111/j.1747-9991.2007.00063.x>.
- Bex-Priestley, Graham (2024): 'Expressivists Should be Reductive Naturalists,' in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 19: Oxford: Oxford University Press.
- Blackburn, Simon (1984): *Spreading the Word*, Oxford: Oxford University Press.
- Blackburn, Simon (1991): 'Just Causes', *Philosophical Studies* 61(½): 3–17. URL = <https://www.jstor.org/stable/4320165>.
- Blackburn, Simon (1998): *Ruling Passions*, Oxford: Oxford University Press.
- Crane, Tim (2001): *Elements of Mind*, Oxford: Oxford University Press.
- Dreier, James (2004): 'Metaethics and the Problem of Creeping Minimalism', *Philosophical Perspectives* 18: 23–44. URL = <https://www.jstor.org/stable/3840926>.
- Dreier, James (2015): 'Explaining the Quasi-Real', in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 10: Oxford: Oxford University Press.
- Enoch, David (2011): *Taking Morality Seriously*, Oxford: Oxford University Press.
- Gibbard, Allan (2003): *Thinking How to Live*, Harvard, Mass.: Harvard University Press.
- Jackson, Frank & Pettit, Philip (1988): 'Program Explanation: A General Perspective', *Analysis* 50(2): 107–117. URL = <https://doi.org/10.1093/analysis/50.2.107>.
- Leary, Stephanie (2021): 'What is Non-Naturalism?', *Ergo* 8: 787–814. URL = <https://doi.org/10.3998/ergo.2253>.
- Oddie, Graham (2005): *Value, Reality, and Desire*, New York: Oxford University Press.
- Raatikainen, Panu (2006): 'Mental Causation, Interventions, and Contrasts,' unpublished manuscript. URL = <https://philpapers.org/rec/RAAMCI>.
- Raatikainen, Panu (2007): 'Mentaalinen kausaatio,' in H. Gylling, I. Niiniluoto & R. Vilkko (eds.): *Syy*, Helsinki: Gaudeamus, 227–237.
- Raatikainen, Panu (2010): 'Causation, Exclusion, and the Special Sciences,' *Erkenntnis* 73: 349–363. URL = <https://doi.org/10.1007/s10670-010-9236-0>.
- Raatikainen, Panu (2013): 'Can the Mental Be Causally Efficacious?', in K. Talmont-Kaminski & M. Milkowski (eds.): *Regarding the Mind, Naturally: Naturalist Approaches to the Sciences of the Mental*. Newcastle: Cambridge Scholars Press.
- Raatikainen, Panu (2018): 'Kim on Causation and Mental Causation', *E-LOGOS - Electronic Journal for Philosophy* 25: 22–47. URL = <https://doi.org/10.18267/j.e-logos.458>.
- Ridge, Michael (2014): *Impassioned Belief*, Oxford: Oxford University Press.
- Rosen, Gideon (2021): "The Modal Status of Moral Principles," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 16: Oxford: Oxford University Press.
- Sinclair, Neil (2012): 'Expressivist Explanations', *Journal of Moral Philosophy* 9: 147–177. URL = <https://doi.org/10.1163/174552412X625745>.
- Sturgeon, Nicholas (1984): 'Moral Explanations', reprinted in G. Sayre-McCord (ed.), *Essays on Moral Realism*, Ithaca: Cornell University Press, 1988.

- Väyrynen, Pekka (2018): 'The Supervenience Challenge to Non-Naturalism', in T. McPherson & D. Plunkett (eds.), *The Routledge Handbook of Metaethics*, New York: Routledge.
- Woodward, James (2003): *Making Things Happen*, New York: Oxford University Press.