

The Hard Problem for Soft Moral Realism

(Forthcoming in *Journal of Philosophy*)

Lei Zhong

The Chinese University of Hong Kong

[Abstract] Several leading moral philosophers have recently proposed a *soft* version of moral realism, according to which moral facts—though it is reasonable to postulate them—cannot metaphysically explain other facts (Dworkin 2011; Parfit 2011; Scanlon 2014). However, soft moral realism is faced with what I call the *Hard Problem*, namely, the problem of how this soft version of moral metaphysics could accommodate moral knowledge. This paper reconstructs three approaches to solving the Hard Problem on behalf of the soft realist: the *autonomy* approach, the *intuitionist* approach, and the *third-factor* approach. I then argue that none of them is successful.

1. Introduction

Moral realism—understood broadly as the view that there are knowable moral facts—is a dominant position in contemporary metaethics. This paper will simply assume that moral realism is correct. But moral realists are deeply divided over the metaphysical nature of moral entities. Moral realism in a *hard* or robust form is committed to the explanatory efficacy of moral facts (see Brink 1989; Railton 1986; Sturgeon 2006; Wedgwood 2007). This is coherent with an influential metaphysical view, according to which any *real* entity must be capable of making a difference to the world—that is, it must be able to metaphysically explain something else. This view can be found as early as in Plato’s *Sophist*, and has been reinforced and developed in contemporary philosophy (see Harman 1977; Hawthorne 2001; Kim 1998; Lewis 1983; Papineau 2016; Shoemaker 1984).

Recently several prestigious moral philosophers, however, have been challenging this metaphysical commitment of moral facts and rather proposing a *soft* version of moral

realism (see e.g. Dworkin 2011; Parfit 2011; Scanlon 2014). In their view, moral entities are *non-natural* entities and hence cannot play explanatory roles in other domains; but owing to the *sui generis* character of morality, it is still reasonable to postulate the domain of moral facts.

Let me introduce the distinction between what I call ‘hard moral realism’ and ‘soft moral realism’ as follows:

[Hard Moral Realism] Moral facts can play explanatory roles in non-moral domains (Boyd 1988; Brink 1989; Railton 1986, 1989; Sayre-McCord 1988; Sturgeon 1985, 2006; Wedgwood 2007).

In contrast,

[Soft Moral Realism] Moral facts are unable to play explanatory roles in non-moral domains (Audi 1997, 2013; Dworkin 2011; Enoch 2010, 2011; Parfit 2011; Scanlon 2014; Setiya 2012; Shafer-Landau 2003; Wielenberg 2010).

The term ‘explanation’ (‘explanatory role’, ‘explanatory power’, etc.) is used in a metaphysical sense.¹ If an event/fact/property E is present *because of* C, I will say that C *explains* E (see Kim 1994; Kment 2014).² Causation is a typical kind of explanation. For example, an assassin’s shooting President Kennedy caused him to die. Then we can say that the assassin’s shooting explains Kennedy’s death. But not all instances of explanation are those of causation. Ontological (or “vertical”) explanations, such as realization and constitution, are clear cases of non-causal explanations. Water’s macro properties, e.g., transparency and solubility, are realized by its microphysical structure. In this case,

¹ I sometimes use the term ‘explanations’ in a secondary sense, referring to *explanatory theories*, which aim to discover explanatory relations in the primary, metaphysical sense.

² My discussion is neutral to what kind of entities (facts, properties, events, etc.) are suitable as explanatory relata. For the sake of convenience, I mainly speak of facts. And I leave it open whether facts should be understood as true propositions or obtaining states of affairs.

water's microphysical structure explains its macro properties, although the former is not the cause of the latter.³

Furthermore, there may also be “horizontal” (or diachronic) explanations that are not causal explanations. For example, mathematical facts may figure in (horizontal) explanations of physical facts, although it's not plausible to say that they *cause* those physical facts (Baker 2005; Batterman 2010; Harman 1965; Lyon 2012; Putnam 1975). Even physical entities may diachronically explain a phenomenon without causing it. According to a number of philosophers, causation is a macroscopic phenomenon that only occurs at the macrophysical level (Norton 2007). If this is true, microphysical events at the fundamental level could not cause anything, although they may still have explanatory powers.⁴ According to other philosophers, on the contrary, only fundamental physical entities can be causally efficacious, whereas higher-level physical entities may play a distinctive explanatory role (Jackson and Pettit 1990).

Some philosophers characterize hard moral realism in terms of causal efficacy (Dworkin 2011; McGrath 2014; Sturgeon 2006). But this is too strong and brings about unnecessary controversies. Considering that moral facts may play explanatory roles in non-moral domains even if they have no causal powers, it is better to formulate hard/soft realism in the light of explanatory efficacy.

In this article, I will argue that soft moral realism is faced with what I call the *Hard Problem*, namely, the problem of how this soft version of moral metaphysics could accommodate moral knowledge. Regardless of whether the *existence* of moral facts is

³ Some argue that abstract facts (such as mathematical, logical, and moral facts) can *constitutively* explain our beliefs about those facts (see Bengson 2015; Cuneo 2018). This is a case of ontological explanation.

⁴ A more plausible picture is that *some* effects have no physical causes, although physical events may be explanatorily sufficient for the occurrences of those effects (see Zhong 2014, forthcoming).

compatible with their explanatory impotence, the *knowledge* of moral facts seems to require that moral facts play explanatory roles in non-moral domains, such as the role of explaining our moral beliefs. If moral facts are *explanatorily isolated* from moral beliefs, then it is mysterious how we can have *epistemic access* to moral facts. Such a requirement also applies to other domains of knowledge, e.g. physics, biology, economics, and even mathematics. According to many philosophers, for a person to know about a domain of facts, those facts must figure in the etiology of her corresponding beliefs. If the real explanations of a domain of beliefs are “debunking” ones—namely, explanations that don’t appeal to the relevant facts—then the beliefs are not knowledge (Field 1989; FitzPatrick 2015; Harman 1977; Joyce 2006; Sosa 2002; Street 2006).

Here let me introduce a particular version of the explanatory condition on moral knowledge:

[E₀] We can have moral knowledge only if moral facts play explanatory roles in non-moral domains (such as the role of explaining moral beliefs).⁵

I regard [E₀] as a *prima facie* plausible thesis; the main purpose of this article is not to provide a positive argument for it.⁶ Rather my paper will focus on whether the soft realist can reasonably reject [E₀] for solving the Hard Problem. Although the worry about whether explanatory inefficacy is compatible with moral knowledge has been raised in the literature (even by soft realists themselves), there are very few serious discussions of it. I will reconstruct three approaches to solving the Hard Problem on behalf of the soft realist: the *autonomy* approach, the *intuitionist* approach, and the *third-factor* approach. Then

⁵ Moral facts may need to explain non-psychological facts for the sake of vindicating moral knowledge (Wright 1992). But for our purposes, we can leave this issue aside.

⁶ [E₀] is even weaker and more plausible than the metaphysical thesis that moral entities can *exist* only if they play explanatory roles in non-moral domains.

I argue that none of them can successfully reject [E₀]. My criticisms of these approaches will partly rely on distinctions/connections between [E₀] and three other explanatory conditions on moral knowledge that are often confused with [E₀].

2. The Autonomy Approach

While hard realism contends that moral facts play explanatory roles in non-moral domains, soft realism denies this. But it is worth noticing that soft realism doesn't necessarily reject the explanatory efficacy of moral facts *as such*; what the soft realist denies is that moral facts can have explanatory efficacy in *non-moral* domains. Even if moral facts fail to explain non-moral facts, one may argue, they can still have explanatory efficacy within the moral realm. This is what I call the 'explanatory autonomy' approach (Parfit 2011; Scanlon 2014; Shafer-Landau 2003).

This strategy is similar to the *causal autonomy* approach in the philosophy of mind. On the autonomy view of mental causation, even if mental properties are incapable of causing fundamental physical properties due to the causal closure of physics (or for some other reason), mental properties can still cause higher-level properties, such as behavioral, social, and biological properties. For the sake of vindicating human agency, mental-physical causation doesn't matter much. As long as mental properties can causally influence higher-level properties, the causal powers of the mind would still be largely preserved, regardless of whether mental properties can cause fundamental physical properties (see Zhong 2011, 2014).

Similarly, it could be contended that even if moral facts can only explain other moral facts, this partial explanatory role of moral entities is sufficient for the metaphysical

and epistemic purposes of moral realism. Russ Shafer-Landau explicitly adopt this strategy:

Why should we insist that the ontological credibility of moral facts depends on their being able to explain things that occur outside of their domain? We don't demand a showing that astronomical facts cause glandular facts, or that harmonic facts cause meteorological facts... If the ontological credentials of such things are impeccable, without their having to causally explain events outside of their own domain, then why should moral facts be held to any higher standard? (Shafer-Landau 2003, p. 104).

The autonomy approach suggests that we don't have to accept [E₀]; we should only accept [E₁], which is compatible with soft realism:

[E₁] We can have moral knowledge only if moral facts play explanatory roles.

However, the autonomy strategy is *not* a promising approach to saving moral knowledge, regardless of whether it is a plausible solution to the problem of mental causation. Here I want to offer two reasons. First, explaining only moral facts is far from sufficient for defending moral knowledge. The same is true of other domains. For example, if we have knowledge about God, it must be the case that God can successfully explain some non-religious facts (say, the phenomenon of design). It is irrelevant whether the existence of God can help explain why there are angels. Shafer-Landau is right that we do not demand that astronomical facts explain glandular facts for the sake of defending astronomical knowledge. But we do require that astronomical facts explain non-astronomical facts, for example, facts about telescope readings and facts about our perceptual experiences.

I will discuss the first point in more detail in later sections. In the current section, let me focus on my second criticism of the autonomy approach. Regardless of whether explaining moral facts is sufficient for moral knowledge, I am afraid that the autonomy approach is unsatisfactory for soft realists' purposes. On closer examination, we will find

that *if* moral facts can explain other moral facts, they must be able to explain non-moral facts as well. See the figure below for illustration:

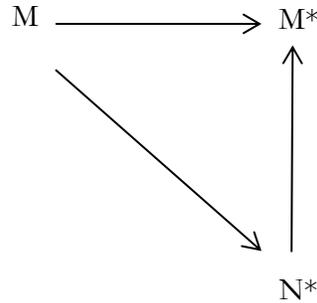


Figure 1

M is a moral property, which is instantiated at time t_1 , and M* is another moral property instantiated at a later time t_2 . M* is realized by a non-moral property N*. It is worth noticing that the M-M* explanatory relation is a diachronic explanation. Certainly, not all cases of moral explanations that the autonomist considers are diachronic explanations. Some moral explanations are synchronic or vertical explanations. Suppose, for example, that the moral fact that we have an overriding moral *reason* to save a drowning baby explains why it is morally *obligatory* for us to do so. This explanation is a case of synchronic explanation. But the autonomist (or anyone who endorses the explanatory efficacy of moral facts) should agree that moral properties can sometimes diachronically explain other moral properties. A person's bad intention explains why what she did is morally blameworthy; some politicians' discriminatory beliefs explain why the policies that they later carried out are unfair; a family's decent education explains why the kids grew up to be good people; and so on. It would be arbitrary to deny diachronic moral explanation while accepting synchronic moral explanation.

In what follows, I will aim to establish that if M diachronically explains M*, *normally* M will explain the realizer of M*, N* (call this conditional the 'Downward

Explanation Principle’).⁷ Let me give an example. Joseph Stalin’s *evilness* explains why there was *injustice* against many intellectuals in the Soviet Union in 1930s. Suppose that the particular way of realizing injustice in this case is forcing those intellectuals to work in labor camps. It seems to follow that Stalin’s evilness also explains why those intellectuals were sent to labor camps. I wish to point out that the downward explanation principle is well supported by the influential difference-making approach to explanation:

[D] *A* explains *B* if *A* makes a difference to whether *B* obtains (Kment 2014; Lewis 1986; Strevens 2004; Woodward 2003).

The difference-making approach is widely used for characterizing causation (Lewis 1973; Pearl 2000; Woodward 2003). Since explanation is a less demanding relationship than causation is, a difference-making approach to explanation is even more plausible than such an approach to causation (see Jackson and Pettit 1990; Strevens 2004; Zhong 2014, 2016). In some cases, even if *A* makes a difference to *B*, *A* does not cause *B*. We can find such examples in both vertical explanations and horizontal explanations. A person’s neurological states make a difference to, and hence explain, her supervenient mental states, but this is not a case of causation. Mathematical facts can explain physical phenomena in a difference-making sense, although mathematical facts don’t cause physical facts.

There are various substantive accounts that can cash out the difference-making idea. Here I will introduce two dominant difference-making frameworks: the (classical) counterfactualist framework (Lewis 1973, 1986) and the interventionist framework (Pearl 2000; Woodward 2003). Consider the counterfactualist framework first.

⁷ I do not have to claim that whenever *M* explains *M**, *M* also explains *N**. To reject soft moral realism, we only need to establish that if *M* explains *M**, *M* explains *N** in *some* cases.

[DC] A explains B in a difference-making sense if B would not have occurred had A not occurred.

According to Lewis’s analysis of counterfactual conditionals, the conditional “if A had not occurred, then B would not have occurred” is understood as “*some* world where neither A nor B occurs, i.e. a $(\sim A \& \sim B)$ -world, is closer to the actual world than any world where A doesn’t occur but B occurs, i.e. a $(\sim A \& B)$ -world” (Lewis 1973). In Figure 1, M explains M^* . Within Lewis’s counterfactualist framework, M^* counterfactually depends upon M ; M makes a difference to whether M^* occurs. That is to say, some $(\sim M \& \sim M^*)$ -world is closer to the actual world than any $(\sim M \& M^*)$ -world. Given that M^* is realized by N^* , any N^* -world is an M^* -world, and any $(\sim M^*)$ -world is a $(\sim N^*)$ -world. Thus, a $(\sim M \& \sim M^*)$ -world is also a $(\sim M \& \sim N^*)$ -world, and $(\sim M \& N^*)$ -worlds are a subset of $(\sim M \& M^*)$ -worlds. So, if it is the case that some $(\sim M \& \sim M^*)$ -world, w , is closer to the actual world than any $(\sim M \& M^*)$ -world, it is also the case that w , as a $(\sim M \& \sim N^*)$ -world, is closer to the actual world than any $(\sim M \& N^*)$ -world. In other words, if M explains M^* , then M will explain N^* —the downward explanation principle is thus established within the counterfactualist framework (see Zhong 2011).

Now let’s turn to the interventionist framework. Roughly speaking, on an interventionist account, x explains y (where x and y are two variables) with respect to a variable set v if an intervention that changes the value of x would also change the value of y when all other relevant variables in v are held fixed at some value (Woodward 2003; Zhong 2014). Interventionism seems to capture an important difference between genuine explanation and mere correlation: if it is a genuine case of explanation, a suitable intervention that changes the explanans would also change the explanandum; if there is

only mere correlation, on the other hand, the explanandum wouldn't change under relevant manipulation of the explanans.

Since in our discussion explanatory relata are focused on facts, we can speak of two values regarding a fact: the obtaining (or *presence*) of a fact and the non-obtaining (or *absence*) of a fact. Hence the value of a fact changes when the fact becomes absent under intervention. The interventionist account of explanation goes as follows:

[D_I] *A* explains *B* in a difference-making sense if an intervention that makes *A* absent would also make *B* absent (while all other relevant variables are held fixed).

In our case, *M* explains *M**. That is to say, if an intervention were to make *M* absent, this intervention would also make *M** absent. Given that *M** is realized by *N**, if *M** is absent, *N** is also absent. It thus follows that if some intervention were to make *M* absent, this intervention would also make *N** absent (while all other relevant variables are held fixed). Therefore, if *M* explains *M**, then *M* will explain *N**—the downward explanation principle is established in the interventionist framework as well.⁸

From what has been argued above, we can see that if moral facts have explanatory powers in the moral domain, they would have explanatory powers in non-moral domains. Thus, if you accept [E₁], you should accept [E₀]. The autonomy approach fails to give any help to soft realism.

3. The Intuitionist Approach

Now consider another soft realist approach. It is worth noticing that every soft realist accepts the doctrine of *moral intuitionism*—here I use the term ‘moral intuitionism’ as a convenient label for the view that (basic) moral knowledge is *a priori*. Intuitionist moral

⁸ For discussion of downward *causation* within the counterfactualist framework, see Zhong (2011).

epistemology is a primary reason that soft realists reject [E₀]: while explanatory efficacy is required for knowledge of natural domains, the explanatory requirement does not apply to the moral domain, which seems to resist empirical investigation (Dworkin 2011; Parfit 2011; Shafer-Landau 2003).⁹ Let me introduce the distinction between moral empiricism and moral intuitionism:

[Moral Empiricism] Moral knowledge is empirical knowledge, in the sense that moral beliefs can only be justified by sensory experience (Boyd 1988; Brink 1989; Railton 1986, 1989, 2005; Sayre-McCord 1988; Sturgeon 1985, 2006).

In contrast,

[Moral Intuitionism] (Basic) moral beliefs can be justified in an *a priori* way, without the mediation of sensory experience (Audi 1997, 2013; Dworkin 2011; Enoch 2010, 2011; FitzPatrick 2008; Huemer 2005; Parfit 2011; Scanlon 2014; Setiya 2012; Shafer-Landau 2003; Wedgwood 2007).¹⁰

In contemporary metaethics, the debate between moral empiricism and intuitionism is sometimes centered on the debate over whether moral knowledge requires moral explanation. For moral empiricists, ethics is continuous with science; we can have moral knowledge only if moral beliefs are results of inference to the best explanation (Brink 1989; Railton 1986, 1989; Sturgeon 1985). By contrast, moral intuitionists deny that explanatory inference is necessary for moral knowledge (Audi 1997, 2013; Enoch 2010, 2011; FitzPatrick 2008, 2015; Parfit 2011; Shafer-Landau 2003).

⁹ Although all soft realists are intuitionists, not all intuitionists are soft realists. A limited number of scholars advocate a combination of hard realism and moral intuitionism (see Cuneo 2006; Wedgwood 2007). This could be an interesting topic for further investigation.

¹⁰ Basic moral beliefs are moral beliefs that don't epistemically depend upon other moral beliefs. Moral intuitionism doesn't have to hold the view that all moral knowledge is immune from empirical evidence; it can maintain only that basic moral knowledge is *a priori*. For example, knowing that torturing a dog for fun is wrong depends upon the empirical knowledge that dogs are capable of feeling pain and pleasure, but we may know, independent of experience, that torturing a sentient being for fun is wrong.

There are two types of empirical knowledge: *perceptual* knowledge and *abductive* knowledge.¹¹ Obviously, we cannot acquire moral knowledge by mere perception. Perceptual experiences, on the standard definition, consist of experiences through the five senses, each of which has a distinctive phenomenal character (Siegel 2016). But no “sensory phenomenal representation” is possible for moral properties (Audi 2013, p. 33). We couldn’t *literally* perceive rightness and wrongness in the same way that we perceive tables, trees, dogs, etc.; a person with perfect perceptual capacities may not detect the moral properties of actions and traits.¹²

Then we are left with the second empirical method: abduction (or *inference to the best explanation*). Inference to the best explanation (IBE) is essential for extending the scope of human knowledge: from knowledge of what is observed to knowledge of what is unobserved or even unobservable, from knowledge of other people’s behavior to knowledge of their mental states, from knowledge about the present to knowledge about the past; and so on (Harman 1965; Lipton 2004). This is exactly why moral empiricists attach importance to explanatory inference (Boyd 1988; Brink 1989; Railton 1986, 1989; Sayre-McCord 1988; Sturgeon 1985). Geoffrey Sayre-McCord puts it this way:

A hypothesis should be believed *only if* the hypothesis plays a role in the best explanation we have of our making the observations that we do... Even if there are moral facts, and even if some of these facts would help to explain others, none will be epistemically accessible unless some help to explain our making some of the observations we do. No matter how perfect the fit between the content of our moral judgments and a moral theory, no matter how stable and satisfying a reflective equilibrium can be established between them, the theory will not gain observational confirmation unless it enters into the best explanation of why some of our observations are made. (Sayre-McCord 1988, pp. 267, 269).

¹¹ Inductive knowledge can be regarded as a particular form of abductive knowledge.

¹² Some philosophers speak of *moral perception* in a loose and metaphorical sense. For them, to say that one perceives a particular moral fact that *P* is just to say that one *directly* knows that *P* without the mediation of inference (see e.g. Audi 2013). For present purposes, it suffices to note that the so-called “moral perception” is a *non-empirical* way of knowing moral facts.

However, according to moral intuitionists, the moral domain at its basic level is immune from empirical investigation (and abduction in particular). Moral justification is ultimately a matter of *a priori* intuition and reflection, rather than a business of explanatory inference (Audi 1997, 2013; Enoch 2010; FitzPatrick 2008; Parfit 2011; Shafer-Landau 2003). As Kieran Setiya says,

Even if ethical facts explain non-ethical effects, as some have urged, it is not by reference to those effects that our beliefs are typically justified. Most often, when we know an ethical fact, we have little idea what it explains, if it explains anything at all (Setiya 2012, pp. 48-49).

Many moral judgments that we make don't seem to rely on any kind of empirical inquiry. Suppose that someone, upon due reflection, believes that killing an innocent person for fun is wrong. But she doesn't have any clue whatsoever how moral facts can help explain natural phenomena. Should we deny, for this reason, that her belief is a justified belief or knowledge? It seems not. Most people form their basic moral beliefs in a way that involves no explanatory inference or empirical investigation at all. So, moral empiricism would lead to an absurd conclusion that most people (perhaps except moral empiricists themselves) have no moral knowledge. Moral empiricism is in conflict with the phenomenology of moral inquiry.

To argue that moral justification is *a priori* is to argue that moral beliefs can be justified without appealing to empirical evidence (regardless of whether *a priori* justification is understood in an internalist or externalist way). *A priori* justification is not queer. We are justified in holding many fundamental truths (epistemic, mathematical, logical, linguistic, etc.) that don't seem to rely on empirical investigation. For example, the IBE principle itself seems to be *a priori*. If we try to justify this principle by appeal to

experience, there would be either circularity or infinite regress (for most recent defenses of the *a priori* of IBE, see e.g. Biggs and Wilson 2016).

I am sympathetic with the possibility of *a priori* moral knowledge, but a full assessment of the debate between moral empiricism and intuitionism is beyond the scope of this paper. Here let us rather focus on a further move made by the soft realist: given the truth of moral intuitionism, [E₀] is false. According to the soft realist, since moral knowledge is not the product of explanatory inference (or empirical inquiry in general), the explanatory efficacy of moral facts is unnecessary for moral knowledge. However, I will argue, this move is problematic. The inference from the truth of moral intuitionism to the falsity of [E₀] is invalid. What moral intuitionism actually challenges is not [E₀], but another thesis [E₂]:

[E₂] We can have moral knowledge only if moral beliefs are results of inference to the best explanation.

While [E₀] is based on the general idea that we can have knowledge about a domain of facts F only if F explains facts in other domains (call it the *Explanatory Tracking Condition on Knowledge*), [E₂] rather derives from the generic thesis that we can have knowledge about a domain of facts F only if the beliefs about F are results of inference to the best explanation (call it the *Explanatory Inference Condition on Knowledge*). The explanatory inference condition is a more restrictive and less plausible account than the explanatory tracking condition is. Whereas the former fails to accommodate some domains of knowledge, the latter can still hold in those domains.

Certainly, the two accounts can converge on some cases of knowledge. Let me give an example. I came back to my apartment. The door is broken; the room is messy; cash and credit cards have disappeared (although it is a little bit comforting that

philosophy books are still there). Then I believe that a burglar broke into my home. Suppose that this is true. My belief should be regarded as knowledge. On the one hand, the burglary metaphysically explains the occurrence of my belief. The explanatory tracking condition applies to this case. On the other hand, my belief was formed as a result of inference to the best explanation. Why do I believe that there was a burglary? This is because assuming a burglary can best explain the phenomena I observed. The explanatory inference condition also works in this case.

However, there are some domains of knowledge that the explanatory inference condition fails to accommodate (but the explanatory tracking condition still can). First of all, knowledge by explanatory inference is *inferential* knowledge. the explanatory inference condition thus cannot accommodate non-derivative, self-justified knowledge. For example, I am aware, by appeal to introspection alone, that I am in pain. My belief about pain is knowledge, although it is not a result of explanatory inference. But pain is still relevant in explaining the occurrence of my belief about pain—it is *pain* that helps to explain why I believe that I am in pain. Whereas non-inferential knowledge is a counterexample to the explanatory inference condition, the explanatory tracking condition can accommodate such a case quite well.

Moreover, since explanatory inference is typically a form of *a posteriori* inference, the explanatory inference condition cannot apply to *a priori* knowledge, such as mathematical knowledge (regardless of whether it is basic or derivative). For example, many believe, solely by mathematical reasoning, that the sum of the angles of a triangle in a plane is equal to 180° . They have no idea whether mathematical propositions can help

explain physical phenomena. Those mathematical beliefs could still be knowledge. Thus, the explanatory inference condition fails in the mathematical domain.¹³

Nevertheless, this is far from saying that mathematical facts cannot explain facts in other domains. Consider an example of mathematical explanation. Honeybees use hexagonal cells to build their honeycombs. But why hexagons? The dominant explanation among biologists is that natural selection would choose those bees that made their honeycombs in the most efficient manner with the minimal amount of wax. This explanation in fact appeals to a mathematical truth, the so-called ‘Honeycomb Theorem’: A hexagonal grid is the most efficient way to divide a Euclidean plane into regions of equal area with least total perimeter (Lyon 2012; see also Baker 2005). Moreover, mathematical facts can figure in the evolutionary explanations of our mathematical beliefs. If our (basic) mathematical beliefs were false—or if putative mathematical facts represented by our beliefs didn’t obtain—then human beings would not have survived and evolved to hold those beliefs (Baker 2005; Batterman 2010; Gibbard 2003; Joyce 2006; Sosa 2002). Hence, the explanatory tracking condition can accommodate mathematical knowledge.

When moral intuitionists deny the explanatory condition on moral knowledge, they seem to have the explanatory inference condition in mind (Audi 1997, 2013; Parfit 2011; Setiya 2012). But I see no good reason why the intuitionist should reject the explanatory tracking condition. As I have mentioned above, the explanatory tracking

¹³ It is worth noticing that what I deny is that explanatory inference is a *necessary* condition on mathematical knowledge. I would leave it open whether mathematical explanation could be *one* epistemic source of mathematical knowledge. According to some philosophers, we are justified in postulating mathematical entities *if* mathematics plays an indispensable role in scientific explanations (Putnam 1975). This is compatible with what I am saying here. If we can have mathematical knowledge either by explanatory inference or by *a priori* intuition, this would be an example of epistemic overdetermination (Shafer-Landau 2003). But even in this case, explanatory inference is an *unnecessary* condition on mathematical knowledge.

condition is true for the mathematical domain, although mathematical knowledge is *a priori*. So, even if the moral domain is *a priori* and non-natural, it doesn't follow that moral knowledge is immune from the explanatory tracking requirement. Thus, [E₀] remains unchallenged.

4. The Third-factor Approach

In this current section, let us turn to an increasingly popular version of soft realism, the so-called 'third-factor' account, which attempts to provide an alternative model of moral knowledge that doesn't assume the explanatory efficacy of moral facts. According to the third-factor account, we can have moral knowledge as long as moral facts and corresponding moral beliefs are reliably caused or explained by a *third factor*, even if moral facts don't help explain the occurrences of moral beliefs (Audi 2013; Enoch 2010, 2011; Wielenberg 2010). See the following figure:

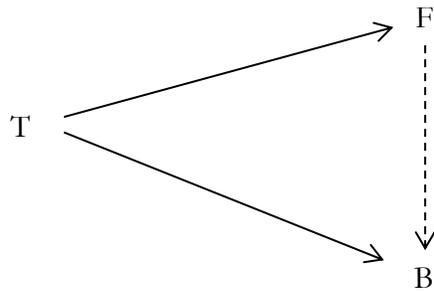


Figure 2

T stands for the third factor, which explains both the belief B and the corresponding fact F. Consider Robert Audi's third-factor account first. For example, I see a person burning a cat for fun, and I come to believe that he is doing something evil. According to Audi, the moral belief that the person is doing something evil (B) is not caused or explained by the corresponding moral fact (F), but is rather caused by the non-moral fact that he is

setting a cat on fire (T), which ontologically explains the moral fact that he is doing something evil (F). Thus, both F and B are explained by the third factor T.

David Enoch proposes a similar but more complicated third-factor account. He puts it this way:

Survival (or whatever) is good; so behaving in ways that promote it is (pro tanto) good; but one efficient way of pushing us in the direction of acting in those ways is by pushing us to believe that it is good to act in those ways. And in fact, as we have just seen, it is good so to act. So the normative beliefs this mechanism pushes us to have will tend to be true (Enoch 2010, p. 431).

People believe, say, that helping others is good. On an evolutionary account, this moral belief (B) is ultimately explained by the non-moral fact that helping behavior contributes to survival (T), but not by the moral fact that helping others is good (F). Moreover, according to Enoch, the non-moral fact T realizes or grounds the moral fact F. That is, T is the third factor that explains both F and B.

Before proceeding to my criticisms of the third-factor account, I want to emphasize a methodological constraint that any reasonable account of moral knowledge is subject to: an account of moral knowledge must not be *ad hoc*. That is, there should be properly analogous third-factor cases in other domains of knowledge. After all, third-factor theorists are not moral quietists; they do not believe that moral epistemology is unique or *sui generis*.¹⁴ The third-factor theorist is, in Mackie's words, "looking for companions in guilt" (Mackie 1977, p. 39).

Now let's consider a typical third-factor example in non-moral domains. An earthquake occurred in Los Angeles in early July. A meteorologist then infers that there will be a tsunami in Los Angeles in late July. Suppose that her belief is knowledge. But

¹⁴ A notorious example of moral quietism is Scanlon (2014).

this belief occurring in early July is *not* explained by the fact that there will be a tsunami in Los Angeles in late July. How could a future event explain a present event? It is safe to exclude the possibility of backward explanation. Rather, both the meteorologist's belief that a tsunami will occur in Los Angeles in late July and the corresponding fact are explained by a third factor, i.e., the fact that an earthquake occurred in Los Angeles in early July (for similar examples, see Goldman 1967; Sosa 1999).¹⁵

On closer examinations, nevertheless, we will find that such third-factor examples fail to raise a serious challenge to [E₀], whereas they call into question another thesis [E₃]:

[E₃] *S* can have moral knowledge that *P* only if the moral fact that *P* explains the occurrence of *S*'s moral belief that *P*.

In some third-factor cases, although *F* doesn't explain the occurrence of *B*, *B* should still be regarded as knowledge. Such cases are thus counterexamples to (a generic version of) [E₃]. However, they don't threaten the plausibility of [E₀]. Consider the tsunami example again. As we know, the fact that there is a tsunami in Los Angeles in late July doesn't explain the meteorologist's corresponding belief occurring in early July. But it is far from saying that tsunamic facts can never explain other facts (including the occurrences of tsunamic beliefs). For example, the fact that there is a tsunami in Los Angeles in late July can explain why some buildings in the coastal area collapsed then. This fact can also explain the occurrences of some tsunamic beliefs—imagine that a person *observed* the tsunami in Los Angeles in late July. On the other hand, although the meteorologist's particular belief that there is a tsunami in Los Angeles in late July is not explained by the

¹⁵ Audi also offers an example of non-moral knowledge. I see a powerful bomb explode next to a soldier but somehow do not directly see the resulting death. Suppose that I believe that the soldier is dead. It is reasonable to regard my belief as knowledge. But, in this case, my belief that the soldier is dead is not explained by the corresponding fact. Rather, the belief that he is dead and the corresponding fact are both explained by a third factor, that is, the fact that a bomb exploded near the soldier (Audi 2013).

corresponding fact, it is reasonable to assume that some other tsunamic facts figure in the etiology of this belief. Probably, this particular tsunamic belief can be causally traced back to other tsunamic beliefs that the meteorologist acquired long time ago, which in turn are explained by the corresponding tsunamic facts.

By the same token, even if there are *some* third-factor cases of moral knowledge in which the moral belief that *P* is not explained by the moral fact that *P*, it doesn't follow that moral facts can never explain moral beliefs. It could be the case that the moral fact explains other moral beliefs, or the case that the moral belief is explained by other moral facts. In either situation, [E₀] still holds. But according to the third-factor moral theorist, no moral facts can explain non-moral facts—*every* instance of moral knowledge is a third-factor case. So, the third-factor account of moral knowledge has no “companions in guilt” in non-moral domains—this account fails to respect the methodological constraint.

Moreover, in what follows, I will develop an argument against the third-factor version of soft realism by proposing a *sophisticated* sensitivity condition on knowledge in contrast with a *simple* sensitivity condition:

[SENSITIVITY] *S* knows that *P* only if the belief that *P* is sensitive to the fact that *P*.
(*The Simple Sensitivity Condition*)

[SENSITIVITY⁺] *S* knows that *P* only if either (i) the belief that *P* is sensitive; or (ii) the belief that *P* is legitimately inferred from other sensitive beliefs. (*The Sophisticated Sensitivity Condition*)¹⁶

S's belief that *P* is sensitive to the fact that *P* if and only if *S* would not believe that *P* if *P* were false—that is, if and only if *S* does not believe that *P* in the closest $\sim P$ -worlds (Nozick 1981; Murphy and Black 2012). There is an intimate connection between sensitivity and

¹⁶ This is similar to Lewis's counterfactual account of causation. Although *c*'s being the cause of *e* doesn't require that *e* counterfactually depend on *c*, it does require that there be a *chain* of counterfactual dependence from *e* to *c*.

explanatory tracking. Suppose that a domain of facts is explanatorily responsible for the occurrences of our beliefs about this domain. It follows from a counterfactual account of explanation that the beliefs counterfactually depend on the corresponding facts—that is, if the facts didn’t obtain, we wouldn’t hold the beliefs. This is equivalent to saying that the beliefs are sensitive to the relevant facts.

It is worth noting that the simple sensitivity condition on knowledge [SENSITIVITY] is an indefensible view (see Sosa 1999). For example, it fails to accommodate *benign* third-factor cases. If e_1 and e_2 are two explananda of a common explanans, e_1 and e_2 do not stand in a relation of counterfactual dependence.¹⁷ Thus, the target belief in the third-factor model doesn’t counterfactually depend on the corresponding fact—in other words, the belief is insensitive. In the tsunami example, the meteorologist’s belief that there is a tsunami in Los Angeles in late July is not sensitive to the corresponding fact. Even if counterfactually no tsunami occurred in Los Angeles in late July, the meteorologist would still believe so based on her knowledge of the earthquake in early July.

In contrast, the sophisticated sensitivity condition [SENSITIVITY+] can accommodate such benign third-factor cases, in which although clause (i) is not met, clause (ii) is satisfied. In the tsunami example, the target belief that there is a tsunami in

¹⁷ When e_1 and e_2 are two effects of a common cause, it follows from the standard *non-backtracking* reading of counterfactuals that no counterfactual dependence holds between e_1 and e_2 (see Lewis 1973). Things become more complicated if the third-factor case involves ontological explanation. Suppose that the moral fact M and the corresponding belief B are explained by a third-factor natural fact N that realizes or grounds M. Does B counterfactually depend on M in this case? My answer is no. We should develop a semantics of counterfactuals that disallows that two explananda of a common explanans stand in a relationship of counterfactual dependence. This could be an interesting topic for further investigation. Here I can only make a tentative suggestion: the reason why B doesn’t counterfactually depend on M is perhaps that in the closest *impossible* worlds where M doesn’t obtain while N obtains, B will still occur. Nowadays many philosophers adopt a non-trivial reading of *counterpossibles* (counterfactual conditionals with impossible antecedents), according to which some counterpossibles are non-vacuously true and others non-vacuously false. See e.g. Bjerring (2014); Brogaard and Salerno (2013); Nolan (1997).

Los Angeles in late July (B) is not a sensitive belief, but it is legitimately inferred from a sensitive belief. This intermediate belief that an earthquake occurred in Los Angeles in early July is sensitive to the corresponding fact. If there were no earthquake in early July, the meteorologist would not believe so. That is, [SENSITIVITY⁺] is satisfied.

[SENSITIVITY⁺] can thus distinguish *malign* third-factor cases from benign third-factor cases. For example, suppose that Descartes’s evil demon has instilled into you a series of *hallucinations* to the extent that your beliefs about the environment are never caused by the external physical objects. Meanwhile, the evil demon brings about the external objects in such a way that your beliefs about your surroundings reliably correlate with the relevant facts. In this case, it is safe to say that your hallucinatory beliefs are not knowledge, even though they are reliably true (Cuneo 2018).¹⁸

Malign third-factor cases such as the evil demon case don’t meet either clause (i) or clause (ii). On the one hand, as opposed to veridical perceptual beliefs, hallucinatory beliefs in the evil demon case are not sensitive to the corresponding facts. On the other hand, unlike the target belief in a benign third-factor case, hallucinatory beliefs are not inferred from other sensitive beliefs—hallucinatory beliefs are basic, non-inferential beliefs in the same way that perceptual beliefs are. Hence, according to [SENSITIVITY⁺], hallucinatory beliefs are not knowledge. See the table below:

	Clause (i)	Clause (ii)
Veridical Perceptual Cases	√	×
Benign Third-factor Cases	×	√
Malign Third-factor Cases	×	×

¹⁸ Reliability is a major selling point of the third-factor account (Enoch 2010). However, as we can learn from the evil demon case, mere reliability is insufficient for knowledge.

Now let us turn back to morality. Is the target moral belief in the third-factor model an inferential belief or not? If yes, we can ask further whether the target moral belief is inferred, at least partially, from other moral beliefs, or instead solely from non-moral beliefs. To answer that the inferential base includes moral beliefs is unsatisfactory. Are the intermediate moral beliefs sensitive to the corresponding moral facts? Does the explanatory condition [E₀] apply to those moral beliefs? The problem is only pushed one step back.

Alternatively, it could be suggested that the target moral belief is inferred from non-moral beliefs alone. In this scenario, we can safely say that (at least some of) those non-moral beliefs are beliefs about the empirical world, such as physical, biological, psychological, and social beliefs—it is unlikely that the non-moral beliefs that could justify a moral belief are solely *a priori*. If moral beliefs are justified in an *a posteriori* way, then this would be a case of moral empiricism. But like other soft realists, third-factor theorists are also committed to an intuitionist moral epistemology.

Here is a reason why soft moral realists should reject moral empiricism. Inference to the best explanation seems to be the only valid form of *a posteriori* inference. If moral knowledge is acquired by empirical inference as we currently assume, moral beliefs must be the results of IBE. In that situation, moral facts would figure in the best explanations of some non-moral phenomena. This almost amounts to saying that moral facts can play (metaphysically) explanatory roles in non-moral domains—in other words, we should accept hard moral realism.¹⁹

¹⁹ I don't hold the view that all empirical knowledge is based on inference to the best explanation. Obviously, *non-inferential* empirical knowledge (such as perceptual knowledge and proprioceptive knowledge) is not a product of IBE. But it is safe to say that *if* moral knowledge is empirical knowledge, it must be inferential empirical knowledge (see Boyd 1988; Railton 1986; Sayre-McCord 1988). I am grateful to a reviewer of the *Journal* for pressing me to address this issue.

It is thus unsurprising that all third-factor theorists hold the view that the target moral beliefs in their models are non-inferential beliefs (Audi 2013; Enoch 2010; Wielenberg 2010). On Audi's third-factor account, we can directly "see" a moral fact without the mediation of inference. For example, although someone's putting a cat on fire caused me to have the moral belief that this person is doing something wrong, my belief is not inferred from any non-moral beliefs.²⁰ Also, in Enoch's version, our basic moral beliefs are clearly non-inferential beliefs, since they were caused by evolutionary forces. However, if no inference is involved in the third-factor model, it looks like a malign third-factor case. Moral beliefs would be similar to hallucinatory beliefs in the evil demon case in that they are neither sensitive nor inferential—[SENSITIVITY⁺] is violated. Therefore, it is unreasonable to regard such moral beliefs as knowledge.

5. Conclusion

As I have argued thus far, the soft moral realist cannot reasonably reject the explanatory condition on moral knowledge [E₀] and hence has trouble accommodating moral knowledge. Certainly, even if the existence of moral knowledge requires that moral facts exert explanatory powers in non-moral domains, it is another issue whether they can *in fact* play such an explanatory role. While some dismiss the viability of moral explanation (Audi 1997; Enoch 2011; Gibbard 2003; Harman 1977; Wright 1992), others attempt to defend the explanatory efficacy of moral facts (Bengson 2015; Cuneo 2006; Railton 1986; Sturgeon 1985; Wedgwood 2007). This is still an ongoing debate, deserving more

²⁰ It is controversial to say that the moral belief in the burning cat example is a non-inferential belief. This moral belief seems to be inferred from the non-moral belief that someone is putting a cat on fire in conjunction with a more basic moral belief, say, the belief that causing pain to an animal is wrong. If so, this is a case in which the target moral belief is inferred partially from another moral belief.

discussion than I am able to provide here. In this article, I have instead argued for a weaker thesis that one should not accept soft realism *if* one is already a moral realist.

Since it is hard to be a soft realist, let's be a hard realist.

References

- Audi, R. (1997) *Moral Knowledge and Ethical Character*, Oxford: Oxford University Press.
- Audi, R. (2013) *Moral Perception*, Princeton: Princeton University Press.
- Baker, A. (2005) "Are there Genuine Mathematical Explanations of Physical Phenomena?" *Mind* 114: 223-238.
- Batterman, R. (2010) "On the Explanatory Role of Mathematics in Empirical Science," *British Journal for the Philosophy of Science* 61: 1-25.
- Bengson, J. (2015) "Grasping the Third Realm," *Oxford Studies in Epistemology* 5: 1-38.
- Biggs, S. and J. Wilson (2017) "The A Priority of Abduction," *Philosophical Studies* 174: 735-758.
- Bjerring, J. (2014) "On Counterpossibles," *Philosophical Studies* 168: 327-353.
- Boyd, R. (1988) "How to Be a Moral Realist?" in G. Sayre-McCord (ed.), *Essays in Moral Realism*, Ithaca: Cornell University Press, 181-228.
- Brink, D. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.
- Brogaard, B. and J. Salerno (2013) "Remarks on Counterpossibles," *Synthese* 190: 639-660.
- Cuneo, T. (2006) "Moral Facts as Configuring Causes," *Pacific Philosophical Quarterly* 87: 141-162.
- Cuneo, T. (2018) "The Evolutionary Challenge to Knowing Moral Reasons," in D. Star (ed.) *The Oxford Handbook of Reasons and Normativity*, Oxford: Oxford University Press, 964-986.
- Dworkin, R. (2011) *Justice for Hedgehogs*, Cambridge: Harvard University Press.
- Enoch, D. (2010) "The Epistemological Challenge to Metanormative Realism," *Philosophical Studies* 148: 413-438.
- Enoch, D. (2011) *Taking Morality Seriously: A Defense of Robust Realism*, Oxford: Oxford University Press.
- Field, H. (1989) *Realism, Mathematics and Modality*, Oxford: Blackwell.
- FitzPatrick, W. (2008) "Robust Ethical Realism, Non-Naturalism and Normativity," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 3, Oxford: Oxford University Press, 159-205.
- FitzPatrick, W. (2015) "Debunking Evolutionary Debunking of Ethical Realism," *Philosophical Studies* 172: 883-904.
- Gibbard, A. (2003) *Thinking How to Live*, Cambridge: Harvard University Press.
- Goldman, A. (1967) "A Causal Theory of Knowing," *Journal of Philosophy* 64: 355-372.
- Harman, G. (1965) "Inference to the Best Explanation," *Philosophical Review* 74: 88-95.
- Harman, G. (1977) *The Nature of Morality: An Introduction to Ethics*, New York: Oxford University Press.

- Hawthorne, J. (2001) "Causal Structuralism," *Philosophical Perspectives* 13: 361-378.
- Huemer, M. (2005) *Ethical Intuitionism*, New York: Palgrave Macmillan.
- Jackson, F. and P. Pettit (1990) "Program Explanation: A General Perspective," *Analysis* 50: 107-117.
- Joyce, R. (2006) *The Evolution of Morality*, Cambridge: MIT Press.
- Kim, J. (1994) "Explanatory Knowledge and Metaphysical Dependence," *Philosophical Issues* 5: 51-69.
- Kim, J. (1998) *Mind in a Physical World*, Cambridge: MIT Press.
- Kment, B. (2014) *Modality and Explanatory Reasoning*, Oxford: Oxford University Press.
- Lewis, D. (1973) "Causation," *Journal of Philosophy* 70: 556-567.
- Lewis, D. (1983) "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61: 343-377.
- Lewis, D. (1986) "Causal Explanation," in D. Lewis, *Philosophical Papers*, vol. 2, Oxford: Oxford University Press, 214-40.
- Lipton, P. (2004) *Inference to the Best Explanation*, 2nd edition, London: Routledge.
- Lyon, A. (2012) "Mathematical Explanations of Empirical Facts, and Mathematical Realism," *Australasian Journal of Philosophy* 90: 559-578.
- Mackie, J. (1977) *Ethics: Inventing Right and Wrong*, New York: Penguin Books.
- McGrath, S. (2014) "Relax? Don't Do It! Why Moral Realism Won't Come Cheap," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 9, Oxford: Oxford University Press, 186-214.
- Murphy, P. and T. Black (2012) "Sensitivity Meets Explanation: An Improved Counterfactual Condition on Knowledge," in K. Becker and T. Black (eds.), *The Sensitivity Principle in Epistemology*, Cambridge: Cambridge University Press, 28-42.
- Nolan, D. (1997) "Impossible Worlds: A Modest Approach," *Notre Dame Journal for Formal Logic* 38: 535-572.
- Norton, J. (2007) "Causation as Folk Science," in H. Price and R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford: Oxford University Press, 11-44.
- Nozick, R. (1981) *Philosophical Explanations*, Cambridge: Harvard University Press.
- Papineau, D. (2016) "Naturalism," in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = <<https://plato.stanford.edu/archives/win2016/entries/naturalism/>>.
- Parfit, D. (2011) *On What Matters*, Oxford: Oxford University Press.
- Pearl, J. (2000) *Causality*, New York: Cambridge University Press.
- Putnam, H. (1975) "What is Mathematical Truth," in *Philosophical Papers*, vol. 1, Cambridge University Press, 60-78.
- Railton, P. (1986) "Moral Realism," *Philosophical Review* 95: 163-207.
- Railton, P. (1989) "Naturalism and Prescriptivity," *Social Philosophy and Policy* 7: 151-174.
- Sayre-McCord, G. (1988) "Moral Theory and Explanatory Impotence," in G. Sayre-McCord (ed.), *Essays in Moral Realism*, Ithaca: Cornell University Press, 256-281.
- Scanlon, T. (2014) *Being Realistic about Reasons*, Oxford: Oxford University Press.
- Setiya, K. (2012) *Knowing Right From Wrong*, Oxford: Oxford University Press.
- Shafer-Landau, R. (2003) *Moral Realism: A Defense*, Oxford: Oxford University Press.
- Shoemaker, S. (1984) *Identity, Cause, and Mind*, Cambridge: Cambridge University Press.
- Siegel, S. (2016) "The Contents of Perception," in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = <<https://plato.stanford.edu/archives/win2016/entries/perception-contents/>>.

- Sosa, E. (1999) "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13: 141-154.
- Sosa, E. (2002) "Reliability and the A Priori," in T. S. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press, 369-384.
- Street, S. (2006) "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109-166.
- Strevens, M. (2004) "The Causal and Unification Approaches to Explanation Unified—Causally," *Noûs* 38: 154-176.
- Sturgeon, N. L. (1985) "Moral Explanations," in D. Copp and D. Zimmerman (eds.), *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*, Totowa: Rowman and Allanheld, 49-78.
- Sturgeon, N. L. (2006) "Ethical Naturalism," in D. Copp (ed.), *The Oxford Handbook of Ethical Theory*, Oxford: Oxford University Press, 91-121.
- Wedgwood, R. (2007) *The Nature of Normativity*, Oxford: Oxford University Press.
- Wielenberg, E. (2010) "On the Evolutionary Debunking of Morality," *Ethics* 120: 441-464.
- Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.
- Wright, C. (1992) *Truth and Objectivity*, Cambridge: Harvard University Press.
- Zhong, L. (2011) "Can Counterfactuals Solve the Exclusion Problem?" *Philosophy and Phenomenological Research* 83: 129-147.
- Zhong, L. (2014) "Sophisticated Exclusion and Sophisticated Causation," *Journal of Philosophy* 111: 341-360.
- Zhong, L. (2016) "Exclusion in Morality," *Grazer Philosophische Studien* 93: 275-290.
- Zhong, L. (forthcoming) "Taking Emergentism Seriously," *Australasian Journal of Philosophy*.