Ethical Emergentism and Moral Causation

**1. Introduction**

Contemporary moral philosophers commonly draw a distinction between naturalist and non-naturalist versions of moral realism. Though both views agree that moral properties are real, objective properties of certain things grounded in certain sets of those things’ lower-level natural properties, naturalist moral realism maintains that these moral properties are natural properties because they are, in some sense, “nothing over and above” those lower-level natural properties that ground them. More specifically, naturalist moral realism may be an analytic or synthetic form of identity-naturalism, which claims that moral properties are identical to—and thus ontologically reducible to—the lower-level natural properties that ground them. According to analytic identity-naturalism, moral properties are ontologically reducible to their grounding natural properties because the concepts of the former are analytically reducible to the concepts of the latter, just as being a bachelor is ontologically reducible to being an unmarried male because the concept of a bachelor is analytically reducible to that of an unmarried male.[[1]](#footnote-1) According to synthetic identity-naturalism, moral properties are indeed ontologically reducible to their grounding natural properties, but the identity relations here are synthetic ones that mimic those between heat and molecular motion, or between water and H2O, where the concepts of the former are not analytically reducible to those of the latter. Alternatively, naturalist moral realism may take the non-reductive form of constitution-naturalism, which rejects the identity of moral properties and their grounding natural properties as well as the analytic reducibility of moral concepts and claims instead that moral properties are constituted by those lower-level natural properties that ground them.[[2]](#footnote-2) Under this theory, which is the analogue of non-reductive physicalism in the philosophy of mind that holds mental properties to be constituted by physical ones that ground them, the relations between moral properties and their grounding natural properties mimic those between tables and the microphysical particles that constitute them.[[3]](#footnote-3) In opposition to these conflicting versions of naturalist moral realism, so-called “non-naturalist” moral realism maintains that moral properties are “non-natural” in the sense of being sui generis and *robustly* irreducible with respect to the lower-level natural properties that ground them: moral properties are different in kind from their grounding natural properties as well as “something over and above” them—they are neither identical to nor constituted by their grounding natural properties.[[4]](#footnote-4) So while this kind of moral realism agrees with constitution-naturalism about the ontological irreducibility of moral properties along with the analytic irreducibility of moral concepts, it further rejects the non-reductive idea that moral properties are constituted by the lower-level natural properties that ground them in favor of an even stronger (or “robust”) kind of ontological non-reductionism.

Standing in contrast to all of these versions of moral realism, however, is a recently developed, novel form of such realism called *Emergentist Ethical Naturalism* that attempts to vindicate both the naturalist idea that moral properties are natural properties and the “non-naturalist” claim that moral properties are sui generis and robustly irreducible.[[5]](#footnote-5) This theory thus attempts to show that those who believe in the kind of robust ontological non-reductionism that typically defines so-called “non-naturalism” can count as ethical naturalists. Additionally, this emergentist theory has some interesting theoretical benefits for moral realists. For example, it demonstrates that emergence can provide a unifying account of several attractive metaethical commitments, where this set of commitments, which I shall explain in the next section, includes the relatively uncontroversial commitments of moral non-deducibility, asymmetrical moral dependence, moral determination, and moral supervenience along with the controversial ones of moral realism, ethical naturalism, robust ontological non-reductionism, and causal efficacy.[[6]](#footnote-6) Furthermore, in contrast to the other forms of moral realism described above, this theory offers some novel, emergentist responses to the supervenience challenges that moral realism faces and, in doing so, allows realists to deal with these challenges while accepting both ethical naturalism—or the claim that moral properties are natural ones—and the “non-naturalist” intuition that moral properties are just too different from the lower-level natural properties that ground them to be identical to or constituted by those properties.[[7]](#footnote-7)

This paper focuses on Emergentist Ethical Naturalism’s commitment to causal efficacy, or the claim that moral properties have causal powers, which carries with it some additional theoretical benefits for moral realists, along with its related commitment to moral causation, which supports its commitment to causal efficacy yet carries with it a battery of strong objections as well as a serious problem that the theory faces. After I reconstruct the theory in sufficient detail, I will discuss these additional theoretical benefits that the theory offers to moral realists in virtue of its commitment to causal efficacy (or causally powerful moral properties). Then, after I locate what I see as the most plausible examples of moral causation—or examples where moral properties seem to be implicated in the causation of things, such as the modestly popular example of the moral properties of things causing our very moral judgments of those things—that support the theory’s commitment to causal efficacy, I will address five strong objections to moral causation, which all threaten to undermine the support offered for the theory’s commitment to causal efficacy. Lastly, I will address a serious problem that Emergentist Ethical Naturalism faces—the problem of “downward moral causation”—that arises once we more fully situate causally efficacious, emergent moral properties into our naturalistic world. Although I will explain this problem in much more detail later on, the basic problem here is that the emergent moral properties that are allegedly responsible for the instances of moral causation offered in support of causally powerful moral properties must be implicated in “downward causation”—or causation from a higher level to a lower one—yet such causation has the theoretically fatal result of causal overdetermination. Adequately responding to this problem is of crucial importance because, if the emergentist-naturalist cannot do so, then the problem remains in force and gives us good reason to reject Emergentist Ethical Naturalism. I will, however, attempt to show that the emergentist-naturalist can adequately respond to the problem and thus that it does not give us good reason to reject the theory.[[8]](#footnote-8)

**2. Emergentist Ethical Naturalism Reconstructed**

Let’s begin by spelling out the theory in more detail. It consists of a bundle of metaethical commitments grounded in (1) a minimalist conception of metaphysical naturalism, (2) a minimalist naturalization assumption, and (3) an emergentist postulate about moral properties. More specifically, the view’s metaethical commitments are grounded in the following:

**Minimal Metaphysical Naturalism**: We live in a naturalistic world, or a world with only natural things, which is just a world with no supernatural or spiritual entities (e.g., God or other ghosts).[[9]](#footnote-9)

**Minimalist Naturalization Assumption:** If something fits into our naturalistic world, then it is natural.[[10]](#footnote-10) And something “fits into” our naturalistic world when it can exist in it and is otherwise continuous with the natural stuff that we know already exists. God, for example, does not fit into our naturalistic world because he, as a supernatural or spiritual being, cannot exist in a world with no supernatural or spiritual entities. But anything that can exist in our naturalistic world and can be coherently combined with the natural stuff that we know already exists and incorporated into a continuous naturalistic reality must also be natural.

**Ethical Emergentism**: Moral properties are emergent properties, where this fact obtains in virtue of a synchronic emergence relation that holds with metaphysical necessity between (a) the moral properties of people, actions, policies, institutions, and any other proper objects of moral evaluation and (b) the lower-level, non-moral properties of these objects that the correct first-order moral theory specifies as the grounds of moral properties.[[11]](#footnote-11)

And the metaethical commitments that these three things ground are:[[12]](#footnote-12)

*Moral Realism*: Moral properties are real, objective properties of actions, people, policies, institutions, and any other proper objects of moral evaluation.

*Robust Ontological Non-reductionism*: Moral properties are sui generis and robustly irreducible with respect to the lower-level, non-moral properties that the correct first-order moral theory specifies as their grounds.

*Causal Efficacy*: Moral properties have causal powers—they can be the causes of certain things or otherwise implicated in the causal, diachronic bringing about of certain things.

*Ethical Naturalism*: Moral properties are natural properties.

*Moral Non-Deducibility*: Propositions about the moral properties of things cannot be analytically deduced from propositions about the lower-level, non-moral properties of those things that the correct first-order moral theory specifies as their grounds (i.e., the so-called “is-ought” gap). We can deduce the former propositions from the latter ones if we have knowledge of the synthetic bridge laws that connect the instantiation of the moral properties to the instantiation of the lower-level ones, but we cannot deduce those former propositions from the latter propositions in virtue of the meaning of these latter propositions or the laws of logic.

*Asymmetrical Moral Dependence*: Moral properties depend for their existence on the lower-level, non-moral properties that the correct first-order moral theory specifies as their grounds, yet these lower-level properties do not depend on moral properties for their existence.

*Moral Determination*: Moral properties are determined by the lower-level, non-moral properties that the correct first-order moral theory specifies as their grounds in the sense that these latter properties synchronically and metaphysically necessitate the moral properties.

*Moral Supervenience*: There can be no difference in moral properties among objects without a corresponding difference in the lower-level, non-moral properties of those objects.[[13]](#footnote-13)

Now let’s take a look at how these metaethical commitments are grounded, which first requires an understanding of what emergent properties are. Let’s begin with a few illustrative examples of emergent properties. A classic example is the wetness of water: individual elements of hydrogen and oxygen are not wet, but when combined or organized in a certain way, they form wet water. In this example the wetness of water emerges from the more basic elements that have only non-wet properties and stand in non-wet organizational relations to each other. Another example is being alive: if certain collections of non-living molecules are combined or organized in certain ways, they form objects that are alive. In this example the object’s property of being alive emerges from the myriad of lower-level properties that characterize the object’s non-living parts and their non-living organizational relations to each other. Yet another example is consciousness: if we combine or organize non-conscious brain cells and non-conscious fluids in certain ways and hook those combinations (i.e., brains) up to other non-conscious, biological objects in certain ways, we get conscious organisms. In this example the organism’s consciousness emerges from the lower-level properties that characterize the organism’s non-conscious parts and their non-conscious organizational relations to each other, particularly the properties that characterize its brain matter.

From these possible examples, then, we can see that emergent properties are properties of wholes that arise from their lower-level properties and that are genuinely novel compared to those lower-level properties.[[14]](#footnote-14) A human organism’s consciousness, for instance, is a property that the whole organism possesses due to an enormous amount of lower-level properties that characterize the non-conscious parts of the organism and their organizational relations to each other (especially brain cells and fluids), which is just to say that this property of consciousness arises from those lower-level properties that, as non-conscious properties, render the property of consciousness novel with respect to those lower-level properties. Digging deeper, emergent properties “arise from” lower-level properties in the sense that they are both determined by and asymmetrically dependent on these lower-level properties,[[15]](#footnote-15) and they are “genuinely novel” compared to the lower-level properties from which they arise in the sense that they are distinct in kind compared to those lower-level properties as well as robustly irreducible to those properties.[[16]](#footnote-16) Consider again our third illustrative example of emergent properties from above. On the one hand, as a property of the whole that is *due to* an enormous amount of lower-level properties that characterize the non-conscious parts of the organism and their organizational relations to each other, it is these lower-level properties that determine the organism’s consciousness by making it conscious rather than not. And while these lower-level properties do not depend on consciousness for their existence (they will depend on something else if anything at all), consciousness depends on them for their existence. On the other hand, the property of consciousness is a new kind of property compared to the lower-level ones from which it emerges—nothing like consciousness occurs at the lower level—and it is neither identical to nor constituted by those lower-level properties, but is rather “something over and above” those properties. On this emergentist picture of consciousness, this property of the whole is not like its mass or organizational structure, which is “nothing over and above” the mass of the lower-level parts or the organizational relations that obtain among them, respectively. On this picture, then, consciousness is not just some organized collection of neurons that performs the functions of consciousness; instead, some collection of functioning neurons is at least partly responsible for the emergence of consciousness—it forms at least part of its emergence base—where consciousness performs its own functions.[[17]](#footnote-17)

Overall, then, emergent properties are to be understood in terms of the following ontological features: they are, on the one hand, determined by and asymmetrically dependent on some set of lower-level properties, and on the other they are sui generis and robustly irreducible with respect to those properties. Yet these are not the only ontological features that define emergent properties. Implicit in the idea of an emergent property is that of a real, objective property: as a genuinely novel property of something, an emergent property must be a real, objective property of that thing that is genuinely novel.[[18]](#footnote-18) Additionally, emergent properties have causal powers,[[19]](#footnote-19) which I take to mean that such properties can either straightforwardly cause certain effects or be otherwise implicated in the causal bringing about of certain effects. So besides being determined by and asymmetrically dependent on, as well as sui generis and robustly irreducible with respect to, some set of lower-level properties, emergent properties are real, objective properties with their own causal powers.

Also, in addition to—and arguably consequent upon[[20]](#footnote-20)—these defining ontological features are some epistemic ones. One is theoretical unpredictability, which refers to the idea that emergent properties cannot, even in principle, be predicted from knowledge of how things operate at the lower level.[[21]](#footnote-21) Another is theoretical inexplicability, which refers to the idea that emergent properties cannot be explained in terms of what goes on at the lower level.[[22]](#footnote-22) These two features must be distinguished from their inductive counterparts, which do not characterize emergent properties.[[23]](#footnote-23) Rather, emergent properties are both inductively predictable and inductively explicable because, respectively, we can predict instantiations of them and explain why they are instantiated if we have knowledge of Nagelian bridge laws that synthetically connect the instantiation of emergent properties with instantiations of certain lower-level properties. If, however, we have no knowledge of such laws, but have to rely instead on our knowledge of the lower level alone, then we will not be able to predict them. And even if we have knowledge of these laws, we can never explain why emergent properties emerge from certain lower-level properties even if we can explain why the former properties are instantiated by citing the latter ones and the former’s emergence from them. Last there is the epistemic feature of theoretical non-deducibility, which refers to the idea that propositions about emergent properties cannot be deduced, even in theory, from a complete knowledge of how things operate at the lower level from which they emerge.[[24]](#footnote-24) Although we can deduce instantiations of emergent properties from instantiations of other properties if we have knowledge of Nagelian bridge laws that connect the former to the latter, we cannot do so if we have only the knowledge of the lower level to rely upon.

Given this characterization of emergent properties, we can now see how Emergentist Ethical Naturalism’s metaethical commitments are grounded. Let’s start with how Ethical Emergentism grounds most of them. Since emergent properties are, by definition, real, objective properties with causal powers, Ethical Emergentism straightforwardly implies casual efficacy by claiming that moral properties are emergent ones. Also, by claiming that (1) moral properties are emergent ones and that (2) this is true in virtue of a synchronic, metaphysically necessary emergence relation obtaining between (a) the moral properties of people, actions, policies, institutions, and any other proper objects of moral evaluation and (b) the lower-level, non-moral properties of these objects that the correct first-order moral theory specifies as the grounds of moral properties, Ethical Emergentism implies moral realism. For (1) implies that moral properties are real, objective properties, while (2) implies that these moral properties belong to people, actions, policies, institutions, and any other proper objects of moral evaluation.

Furthermore, by claiming that moral properties are emergent ones because they emerge synchronically and with metaphysical necessity from the lower-level, non-moral properties that the correct first-order moral theory specifies as their grounds, Ethical Emergentism implies asymmetrical moral dependence, robust ontological non-reductionism, and moral non-deducibility. For on the one hand, since moral properties emerge from these lower-level properties, they must be asymmetrically dependent on them as well as sui generis and robustly irreducible with respect to them. On the other hand, by emerging from these lower-level properties, propositions about moral properties cannot be theoretically deduced from those pertaining to lower-level properties, which is just to say that propositions about the moral properties of things cannot be analytically deduced from propositions about their lower-level properties in virtue of meaning or the laws of logic. Also, by postulating a synchronic, metaphysically necessary emergence relation between moral properties and the lower-level, non-moral properties that the correct first-order moral theory specifies as their grounds, Ethical Emergentism implies moral determination: since moral properties emerge synchronically and with metaphysical necessity from those lower-level properties, it follows that those lower-level properties determine the moral ones synchronically and with metaphysical necessity, which is just to say that those lower-level properties metaphysically necessitate the moral ones synchronically.

Next we have moral supervenience, which Ethical Emergentism grounds as follows. By postulating that moral properties emerge synchronically and with metaphysical necessity from whatever sets of lower-level, non-moral properties that the correct first-order moral theory specifies as their emergence bases, Ethical Emergentism implies that these latter sets of properties determine moral properties synchronically and with metaphysical necessity. However, if these latter sets of properties determine moral properties in this fashion, then there can never be a moral difference among objects without a corresponding lower-level difference among them. For suppose that some object, O, possesses some moral property, M, that emerges from, and is thus determined by, some set of O’s lower-level, non-moral properties, N, both synchronically and with metaphysical necessity. Since N metaphysically necessitates M, anything with N must also have M. And because N necessitates M synchronically, anything with N must also have M at the same time; there can be no lag between the instantiation of N and that of M. Yet if anything with N must also have M at all times during which it has N, then there can be no M-difference among objects without a corresponding N-difference among them.[[25]](#footnote-25)

Last there is ethical naturalism, which is grounded as follows. Assume first Minimal Metaphysical Naturalism: we live in a naturalistic world, which is one without supernatural or spiritual entities; all objects, properties, and so on—including all proper objects of moral evaluation and their non-moral properties that, according to the correct first-order moral theory, ground moral properties—are natural. Then we posit Ethical Emergentism, which claims that moral properties emerge synchronically and with metaphysical necessity from the lower-level, non-moral properties of proper objects of moral evaluation that the correct first-order moral theory specifies as their grounds. This grounds moral properties exclusively in natural properties and has them necessarily arise from these grounding properties alone, without any outside aid, in *any* world, including our actual naturalistic one, in which these grounding properties happen to be instantiated, which fits moral properties into our naturalistic world and thereby naturalizes them according to the Minimalist Naturalization Assumption.

**3. The Benefits of Causal Efficacy**

Besides its ability to unify those eight metaethical commitments and to explain moral supervenience while affirming both the naturalness and the robust irreducibility of moral properties, Emergentist Ethical Naturalism offers some additional theoretical benefits to moral realists in virtue of its commitment to causally powerful moral properties. First of all, by implying that moral properties have causal powers, Ethical Emergentism implies that moral properties satisfy a rather demanding ontological principle requiring real, genuinely existing entities to have causal powers. Applied to moral properties, this principle maintains that moral properties are real only if they have causal powers. Though defenders of moral realism are free to reject the validity of this principle—as will realist-epiphenomenalists who believe that moral properties are real ones with no causal powers—the emergentist-naturalist can remain safely non-committal about it and maintain that, even if it is valid, moral properties satisfy it because, as emergent properties, they indeed have their own causal powers.

Furthermore, consider the following anti-realist syllogism based on this ontological principle and an epiphenomenalist premise:

(1) Moral properties are real only if they have causal powers.

(2) Moral properties have no causal powers.

(3) Therefore, moral properties are not real.

Since moral properties satisfy the ontological principle given Ethical Emergentism, Emergentist Ethical Naturalism allows moral realists to respond to this argument without having to choose between rejecting the principle or denying the epiphenomenalist premise by resorting to a standard form of naturalistic realism that construes moral properties as “nothing over and above” the lower-level, causally powerful, natural properties that ground them. For if moral properties satisfy the principle because Ethical Emergentism is true, then the realist can respond to this argument by denying the epiphenomenalist premise, which is now false given Ethical Emergentism, rather than the principle. And, since Ethical Emergentism warrants rejecting the epiphenomenalist premise, realists need not resort to the idea that moral properties are “nothing over and above” their natural, causally powerful grounding properties to reject it.

Finally, by implying that moral properties have causal powers, Ethical Emergentism delivers a crucial part of what seems to be ontologically required for perceptual moral knowledge, which is knowledge that may be important for ethical theorizing by supplying singular ethical judgments for reflective equilibrium.[[26]](#footnote-26) Such knowledge at least requires (a) the perception of moral properties along with (b) the epistemic equipment for acquiring such knowledge. It is also quite plausible to suppose that the perception of moral properties—and thus perceptual moral knowledge—requires a causal connection between the moral properties and our epistemic equipment such that the moral properties causally affect this equipment and thereby produce the perceptions of these properties.[[27]](#footnote-27) If perceptual moral knowledge requires the perception of moral properties, and such perception requires that these properties be able to causally affect us so as to be able to produce these perceptions, then it follows that perceptual moral knowledge requires causally powerful moral properties. Now Ethical Emergentism does not vindicate the idea that we have the requisite epistemic machinery for perceptual moral knowledge, nor does it answer the perplexing question as to how moral properties causally affect our epistemic equipment and thereby produce moral perceptions.[[28]](#footnote-28) Nevertheless, by implying that moral properties have causal powers, Ethical Emergentism delivers the causally powerful moral properties that seem to be required for perceptual moral knowledge, and by doing so it gets the possibility of such knowledge off the ground.

**4. Examples of Moral Causation**

While delivering causally powerful moral properties may offer these theoretical benefits to moral realists, this will not matter if the commitment to such properties turns out to be unfounded or unjustified. Do we have good reason to believe that moral properties have causal powers? Can we cite any actual instances of moral causation to warrant thinking that they have causal powers? Any answers here will of course be controversial, and I cannot pretend to offer examples of moral causation that are as persuasive as the bountiful examples of mental causation, but I still think that the idea that moral properties have causal powers is at least a presumptively reasonable one in light of supporting examples where moral properties appear to be causally implicated in the bringing about of other things.[[29]](#footnote-29)

Let’s first consider some examples of moral causation that other commentators have offered. The first are those modestly popular ones that I alluded to earlier in which the moral properties of things appear to be causal factors behind our moral judgments of those very things.[[30]](#footnote-30) So for instance, why do we immediately judge Gilbert Harman’s hoodlums to be wrong in setting that poor cat on fire? Or why do most of us judge Hitler to be evil and his actions against Jews and other groups to be wrong? A classic, presumptively reasonable answer from Nicholas Sturgeon is that we make these judgments about Hitler and his actions because, in fact, Hitler was evil and his actions were wrong.[[31]](#footnote-31) Now we can give two slightly different causal interpretations of Sturgeon’s answer here, and either way the example supports the commitment to causally powerful moral properties.[[32]](#footnote-32) We could, on the one hand, understand the moral properties as straightforward causes: the moral properties of what those hoodlums are doing and of Hitler and his actions cause us, at least in part, to make these judgments about them. On the other hand, we could understand the relevant objects or events as the causes of these judgments, but their moral properties are implicated in their causation of these judgments. To fully understand this possibility we must turn to Ted Honderich’s discussion of causally relevant properties and his example of a pear weighing down a scale, which it does in virtue of its weight as opposed to, say, its color.[[33]](#footnote-33) Here the pear—the object—causes the scale to go down, and it does so in virtue of certain properties rather than others. These properties are causally relevant factors here—they are clearly implicated in the causal bringing about of the scale going down—but the cause is, strictly speaking, the object. We can say the same thing if we understand the cause as an event rather than an object: the event of you putting the pear on the scale, rather than the pear itself, causes the scale to go down because of the pear’s weight. But either way we understand it, we have a case where a property is directly implicated in the causation of something without being the cause. Similarly, then, the examples on offer here might be ones in which moral properties are directly implicated in the causation of our moral judgments even though they are not, strictly speaking, the causes of these judgments. Perhaps it is actions, or the events of our performing them, that cause moral judgments of them in virtue of their moral properties. Let’s call these cases where our moral judgments are the explananda “judgment-cases.”

Other examples from other commentators cite our actions or our motivations, rather than our judgments, as the explananda. One rather popular example here is that the moral properties of society can partly cause social unrest.[[34]](#footnote-34) One might cite, for instance, the wrongness of the Trump administration’s policy of separating immigrant families at our borders or of the excessive violence that black people sometimes face at the hands of police officers as at least part of the cause as to why there has been so much protest against this policy and this violence. Alternatively, one might cite the wrongness of these things as the causally relevant property that is implicated in the causation of these protests by these things, which are, strictly speaking, the causes of the protesting (i.e., the wrongness is not the cause, but the relevant policy and events do the causing in virtue of their wrongness). It does not really matter how we understand these examples, however, because they, at best, collapse into judgment-cases. For if we assume that those protesting against the Trump administration’s separation policy or excessive police brutality are right to oppose these things, then given the plausible assumption that it is wrong to oppose those who are right, the opposition to those protestors is wrong in opposing the protestors. Moral properties, then, cannot be implicated in the bringing about of this opposition (it is not a response to the rightness of the original protesting), so we must instead explain it in terms of the original protesting, the opposition’s negative moral judgments of the original protesting, and other psychological factors, such as a desire strong enough to motivate their counter-protesting. However, if we must do this for the counter-protestors, then we should similarly explain the original protestors’ behavior by citing the relevant policy and events, their negative moral judgments of those things, and other psychological factors, such as sufficiently strong desire that motivates their protesting.[[35]](#footnote-35) But then the only way to bring moral properties back into the picture as being causally implicated in the bringing about of social unrest would be for them to be causally implicated in the making of the relevant moral judgments that are, in turn, partly responsible for the unrest. This would show that these new examples are just more judgment-cases in disguise, and so these “social-unrest-cases” will not be treated here as fresh, supporting examples of moral causation.[[36]](#footnote-36)

Two other examples that cite our actions or motivations as the explananda are the following. One I attribute to Nicholas Sturgeon, and it cites Hitler’s evil as a causal factor behind his morally horrific actions and practices toward other people.[[37]](#footnote-37) The other comes from Terence Cuneo, who invites us to consider cases where a person’s virtue appears to be the “configuring cause” of their being motivated appropriately.[[38]](#footnote-38) Unfortunately, these cases seem to be best understood as cases of psychological causation rather than moral causation. Consider first the case of Hitler’s morally horrific actions. These are probably best explained as the expression of non-moral, psychological character traits that an anti-realist could believe in, such as his ruthlessness or his disregard for human life, rather than the moral qualities to which these non-moral qualities give rise. In other words: it was not Hitler’s evil that caused his behaviors, but rather the non-moral, psychological qualities that made him evil. Something similar seems true if we consider someone’s virtue rather than their being evil. It is not the moral quality of moral excellence (i.e. virtue) that causes one to be motivated to act appropriately, but rather the non-moral quality that is virtuous or that makes one virtuous to the extent that it is present. In Cuneo’s example, it is Sam’s compassion that makes him motivated appropriately, where his compassion is cashed out as a virtue that is constituted by a set of counterfactually stable, psychological dispositions. Though being compassionate is indeed a virtue such that this quality is itself morally excellent or at least confers moral excellence on its possessor to the extent that it is present, it seems more plausible to hold that it is the compassion—the constellation of psychological dispositions that anti-realists could surely believe in—causing the appropriate motivation rather than the moral excellence that this compassion grounds. In any event, I will not treat these examples that cite our actions or our motivations as the explananda as genuine examples of moral causation.[[39]](#footnote-39)

At this point, then, it looks like the commitment to causal efficacy rests entirely on the judgment-cases. However, I would like to offer what I shall call “compensation-cases” as further, provisional support for this commitment. Consider the fact that, when we wrongfully victimize other people in certain ways, we thereby incur compensatory duties toward them, which means that we perform obligatory actions by providing appropriate compensation to our victims but act wrongly if we do not. Also, if we had not wronged these other people in the first place, then we would neither act wrongly by failing to do whatever it is that we must do to appropriately compensate them nor obligatorily by doing so. Without the previous wrongdoings, the same actions do not have the same moral statuses as they would have if we had committed those wrongdoings. So, by way of illustration, consider the hilarious scene from *The Big Lebowski* where John Goodman’s character, Walter Sobchak, takes a tire iron and destroys some other guy’s car in a fit of rage. In this case, since Walter lacks moral justification for what he is doing, he owes the car’s owner whatever sum it costs to repair or replace the car, and so he performs an obligatory action by paying this sum, whereas he acts wrongly by failing to do so.[[40]](#footnote-40) However, if Walter had not destroyed the car, then it would neither have been obligatory to give the car’s owner the same sum of money nor wrong to fail to do so. In this and other similar cases where we wrong others and thereby owe them compensation, it seems that the wrongness of our actions has effects on the moral statuses of future actions or omissions.

These cases may strike some as odd and unintuitive, while others may be skeptical that we have genuine cases of moral *causation* here instead of, at most, some kind of non-causal determination. Though these are valid concerns, I do not think that they are enough to conclusively undermine these compensation-cases as examples of moral causation. For even if they are odd and unintuitive, there are two considerations that strongly support their legitimacy. First of all, they seem to be cases where wrongness *functions as a difference-maker*: the presence of wrongness makes a difference because without the previous wrongdoings, the same actions do not have the same moral statuses as they have when previous wrongdoing has been committed. Second, the determination going on in them seems to count as moral causation because the previous wrongness in these cases is clearly implicated in the causal, diachronic bringing about of actions or omissions with certain moral properties. While the previous wrongness does not bring about these actions or omissions by themselves, they appear to work in tandem with human agency to do so: the actions or omissions flow from human agency, while their moral statuses are made what they are by the presence or absence of previous wrongdoing. We therefore have cases here where moral wrongness functions as a difference-maker that is clearly implicated in the diachronic bringing about of certain actions or omissions with certain moral properties, which renders such wrongness sufficiently similar to a causal factor to warrant treating it, at least provisionally, as such a factor.

**5. Objections to Moral Causation**

Thus far I have tried to justify Emergentist Ethical Naturalism’s commitment to causally powerful moral properties by offering some examples of actual moral causation wherein moral properties are causally implicated in the diachronic bringing about of things and thereby demonstrate their causal powers. Underlying the theory’s commitment to causal efficacy, then, is the further commitment to the reality of moral causation, and so before we can accept the former commitment in virtue of the latter, some very potent objections to believing in moral causation must be adequately met, as otherwise we would not be able to accept the reality of moral causation in the first place and thus would not be able to accept causally powerful moral properties in virtue of such causation. Due to space limitations and the fact that defending the reality of moral causation is not my central concern here, however, I shall consider only five objections to such causation.

The first objection is that we are unjustified in believing in moral causation because we do not need to implicate moral properties in the causation of things such as our moral judgments.[[41]](#footnote-41) To return to the Hitler example, we do not need to postulate actual evil or wrongness to causally explain our moral judgments about him being evil or his actions being morally wrong. All we really need to postulate here is (1) some set of beliefs on our part that take us from the non-moral facts to moral conclusions along with (2) the non-moral facts. So, for example, in order to explain why we evaluate Hitler’s genocide of Jews and other groups as morally wrong, we only need to postulate that we hold sets of beliefs that include the moral belief that genocide is wrong along with the actual genocide. In order to explain why we evaluate Hitler as an evil person, we only need to postulate that Hitler did certain things along with our beliefs that people who do such things are evil.

This first objection, however, does not sink the commitment to moral causation. First of all, even if we do not need to postulate moral properties or facts to causally explain psychological or social facts, the compensation-cases remain unscathed. Furthermore, the premise that we are unjustified in believing in moral causation because we do not need moral properties to causally explain things such as our moral judgments seems false. We do not, for example, need to postulate an exploding singularity to causally explain the universe’s existence; we can instead explain it by saying that God created it *ex nihilo*. We also do not need God to causally explain the universe’s existence; we can explain it, instead, by recourse to some other set of supernatural beings or to an exploding singularity. The fact that we do not *need* an exploding singularity or God to causally explain the universe, however, does not imply that we are unjustified in believing in either explanation. We are justified in believing in the exploding-singularity explanation if the total evidence points toward it; and even though we are unjustified, in my estimation, in believing in the theistic explanation, this is not because God is not needed to explain the universe’s existence, but rather because there is (a) no good evidence for this explanation and (b) good evidence against it (e.g., the amount and severity of the world’s evil).

A second objection to the commitment to moral causation maintains that such causation is suspect because causation requires a mechanism, yet it is not easy to see what this would—or even could—be in the case of moral causation.[[42]](#footnote-42) But this objection does not succeed for two reasons. First, causation may not require a mechanism.[[43]](#footnote-43) Second, even if moral causation does require a mechanism, the fact that we cannot see what it would or could be does not sink the commitment to moral causation if we already have good evidence for it. Charles Darwin had good evidence for his theory of descent with modification, and so his inability to discern the mechanism by which this modification occurred only indicated that this mechanism was a mystery. Similarly, since the judgment-cases and compensation-cases constitute good evidence for moral causation, our inability to discern its mechanism (assuming that there is one) only indicates that this mechanism is a mystery; it does not indicate that there may be no mechanism, and thus no causation, in the first place.

The last three objections to moral causation are based in what are taken to be theoretical virtues. The basic complaint of each of these objections is that we are unjustified in believing in moral causation because it is not indispensably involved in, but instead competes with, the best explanation of our immediate moral judgments or intuitions, such as the immediate judgment “that is so wrong!” that we would have when rounding the corner and seeing Harman’s hoodlums lighting that poor cat on fire, where this best explanation is an anti-realist one that better exhibits some theoretical virtue.[[44]](#footnote-44) The first of these three objections appeals to considerations of mystery, and it runs as follows. Realist explanations that postulate moral properties as causal factors behind moral judgments will have to grant that the mechanism by which this moral causation between the moral properties and the judgments occurs is mysterious. By contrast, anti-realist explanations will not be similarly plagued by a mysterious causal mechanism that connects the explanans and the explanandum because it only cites non-moral properties or facts to explain the relevant judgments. For example, the anti-realist could explain our moral intuition that Harman’s hoodlums are acting wrongly by citing (1) the non-moral fact pertaining to what the children are doing, (2) our epistemic apparatus that allows us to detect this fact, (3) a background moral belief that setting cats on fire is wrong, where holding this belief is explained as part of a mere—and ultimately mistaken—social convention that some human or group of humans invented in the past due to their repugnance of burning cats alive, and (4) the ability to immediately judge that what the hoodlums are doing is wrong given this background belief and the non-moral fact detected. Since this anti-realist explanation is committed to no mysterious causal mechanism connecting the explanans to the explanandum, it is less mysterious, and thus better, than the realist one, and so we are unjustified in believing in the inferior realist explanation that involves moral causation.[[45]](#footnote-45)

The last two objections, by contrast, appeal to considerations other than mystery to ground the anti-realist’s claim of explanatory superiority. One appeals to *parsimony*: the above anti-realist explanation of our intuition about the hoodlums is simpler than the realist one because it explains the intuition without positing moral properties, and since it is simpler, it is better. The other is based in *consilience*: the above anti-realist explanation only cites non-moral properties or facts, such as psychological ones, which can explain more classes of facts than moral properties or facts can, and so the anti-realist explanation is better.[[46]](#footnote-46)

Although these objections are quite forceful, they do not sink the commitment to moral causation. For starters, they do not undermine the compensation-cases. They are also rather dubious in certain places. It is not immediately clear, for instance, that there is less mystery involved in the anti-realist’s explanation. I for one cannot explain how our epistemic equipment can detect the non-moral fact about what the children are doing, and when it comes to the mechanism connecting our detection of this fact and the background moral belief to the moral intuition that those deplorable hoodlums are acting wrongly, I can only lamely say, “our faculty of reason.” Of course, a cognitive psychologist or neuroscientist might be able to explain our detection of the relevant non-moral fact, and perhaps it is enough to cite our faculty of reason as the other mechanism to avoid the infection of mystery. If so, however, then the realist can likewise cite our faculty of reason as the mechanism that connects moral properties to our detection of them, which would mean that the anti-realist explanation is not less mysterious after all.

Furthermore, and perhaps more importantly, even if we grant that the anti-realist explanation is less mysterious, simpler, or more consilient, it does not follow that it is better. The better explanation here, in my estimation, is the one that is more likely to be true, yet an explanation’s being less mysterious just means that it is more intelligible to us; it does not mean that it has a greater likelihood of being true. Anthropomorphic explanations of things, such as a theistic explanation of the universe or other supernatural explanations of apparent natural events (e.g., floods), might be less mysterious and thus more intelligible than alternative, non-anthropomorphic explanations that cite no human-like being as responsible for the explananda, but even so this would not give us good reason to think that the anthropomorphic explanations are more likely to be true. Conspiracy theories of social events (e.g., the COVID-19 pandemic) might be less mysterious and thus more intelligible than alternative explanations that cite no conspiracies, but once again, this greater intelligibility due to less overall mystery would not give us good reason to think that the conspiracy theories are more likely to be true. Similarly, an explanation, E1, may instead be simpler or more consilient than another, E2, but neither gives us a good reason to think that E1 has a greater shot at being true. One could hold a very consilient theory, for instance, by believing that God’s will explains the universe’s creation and its persistence through time as well as every event that happens in the universe and every feature that characterizes it, yet the mere consilience of this theory, which is quite large, does not at all suggest that it is true or more likely than naturalistic explanations of these things. One could also hold a Berkeleyan theory of reality according to which the universe and its events and its features—including us—are only ideas contained in the one thing that exists: God’s infinite mind. Insofar as it posits only one actual thing to explain what we experience as reality, this theory seems about as simple as they come, yet its supremacy here does not at all suggest that this theory is true or more likely than a naturalistic theory of reality that posits more stuff and rejects the supernatural altogether. The anti-realist explanation here might be less mysterious, simpler, or more consilient than the realist one, but none of these render it the better explanation because none of them indicate a greater likelihood of being true.

Finally, these objections seem to assume that the intuitions to be explained are those of people who already reject violence against cats, which is probably due to the fact that the authors and the audiences of these discussions already have—rightfully so—negative attitudes to violence against cats and are horrified by Harman’s example (I am definitely horrified by it!). But what if the person rounding the corner previously despised cats as dangerous and worthless vermin that can be mistreated without limitation? What if their experience is a conversion experience wherein the person ultimately changes their tune on how cats can be treated? In that case it is not so obvious that the anti-realist can explain these strange new intuitions by recourse to the person’s previous moral beliefs. They certainly could not cite any beliefs about it being wrong to burn cats alive because the person in question had beliefs that conflicted with this belief; other beliefs or something other than such beliefs would have to fulfill the anti-realist’s explanatory need here. In any event, in this slightly modified case, it seems a bit more plausible to cite the wrongness of the action as causing—or as being causally implicated in the bringing about of—these strange new intuitions about it being wrong to burn cats alive that contradict previously held beliefs about cats.

Although they all initially seem quite powerful, none of these objections that I have considered here clearly defeats the commitment to moral causation; at most, it looks like these objections invite further debate. Since this commitment is not clearly defeated by any of the five objections considered, we can provisionally conclude that Emergentist Ethical Naturalism’s commitment to causally powerful moral properties is still justified by actual moral causation and move on to our final issue with the theory: its problem with downward moral causation.

**6. The Problem of Downward Moral Causation**

The theory’s problem with downward moral causation stems from its postulation of emergent moral properties as those responsible for the actual instances of moral causation that it relies upon as support for its commitment to causal efficacy. To see the problem here, however, we must first fully situate these properties into our naturalistic world. So far, the theory has only postulated that moral properties emerge from the lower-level, non-moral, natural properties that, according to the correct first-order moral theory, ground moral ones. However, these lower-level natural properties that are supposed to ground moral ones are not themselves fundamental properties—they, too, will be grounded in lower-level properties. In fact, these lower-level natural properties will ultimately be grounded in some enormously complex set of fundamental physical properties, or properties of the fundamental physical stuff that ultimately constitutes all proper objects of moral evaluation, and so moral properties themselves will ultimately be grounded in fundamental physical properties.[[47]](#footnote-47) Given this, we can then re-purpose Jaegwon Kim’s arguments against mental emergentism to apply to ethical emergentism.[[48]](#footnote-48)

The first part of the argument here tries to establish that actual causal relations that involve emergent moral properties as causal factors must also involve downward moral causation. To see how, let’s begin by supposing, as Emergentist Ethical Naturalism does, that there are emergent moral properties that are causal factors behind some of the causation that occurs in the world. For any of these emergent moral properties, M, it is the case that it is grounded in some set of lower-level natural properties, N, as specified by the correct first-order moral theory. Now, M has, by hypothesis, causal effects in the world, and there are only three possibilities for what these effects will be: lower-level effects, same-level effects, or higher-level effects. If M has lower-level effects, then this is already downward moral causation. If, however, M has same-level effects, which we can call M\*, then since M\* will likewise be grounded in some set of lower-level natural properties, N\*, it seems that the only way for M to cause M\* will be by causing N\*, and so again we have downward moral causation. Finally, if M has higher-level effects, which we can call H, then since H will also be grounded in some set of lower-level properties, L, it seems that the only way for M to cause H will be by causing L. But since L is at the same level as M, higher-level causation entails same-level causation. And since same-level causation entails downward causation, higher-level causation does so as well. So, no matter what, it seems that emergent moral properties that figure in actual causal relations must be responsible for downward moral causation.

The second part of the argument then purports to show why such downward causation is problematic. Recall first that, for any given emergent moral property, M, that by hypothesis figures in actual causal relations, it is grounded in some set of lower-level natural properties, N, as specified by the correct first-order moral theory. Next recall that, when we further situate M into the world, we see that it will ultimately be grounded in some enormously complex set of physical properties, P, because N will itself ultimately be grounded in P. Now we just saw, in the previous paragraph, that M will be responsible for downward moral causation—it will have lower-level effects. Yet there are only two possibilities for what these lower-level effects could be: they are either (1) already physical effects at the lowest level, or else (2) they are effects at some higher level where the properties here are ultimately grounded in some set of physical properties at the lowest level. Accordingly, M will have physical effects: it either (1’) straightforwardly has such effects, or else it (2’) has such effects because, in order to have effects at some higher level than the physical, it must cause the set of physical properties that ground the properties at that higher level. However, if M must have physical effects, then given the Causal Closure of the Physical Principle, which maintains that every physical effect is sufficiently caused by some set of physical properties, it follows that any physical effect of M will also have an additional, physical cause, P, where P is the lowest-level emergence base of M itself. And since M is, by hypothesis, an emergent property, which makes it robustly irreducible with respect to anything from which it emerges, it follows that any physical effect of M will have two distinct sufficient causes: M and P. Yet this means that these physical effects of emergent moral properties are causally overdetermined, which is a bad consequence of downward moral causation that warrants the rejection of such causation and thus of Emergentist Ethical Naturalism as a whole.[[49]](#footnote-49)

Although this argument brings up difficult issues surrounding causation that call for more extensive treatment than can be given here, the emergentist-naturalist can still blunt the argument’s force by putting pressure on (1) the claim that causally efficacious, emergent moral properties must be responsible for downward moral causation, (2) the derivation of overdetermination from such downward moral causation, or (3) the assumption that overdetermination would be a bad result that warrants rejecting such causation and therefore the theory as a whole. Starting with the first claim here, it is questionable because the argument for it rests crucially on the Kimian assumption that the only way for causally efficacious emergent properties to cause other properties is for them to first cause the supervenience bases of these other properties. However, the compensation-cases I offered earlier suggest that moral properties can be causal factors behind other moral properties without first causing the instantiation of their supervenience bases. For in those cases, it is *the moral agents*, rather than the moral properties of their previous wrongdoings, that instantiate the supervenience bases of the moral properties—it is the moral agents that choose to provide compensation or not. But it is the previous wrongdoings—and in particular, the wrongness of those doings—that determines the moral status of whatever the moral agents happen to choose later on. This suggests, contrary to the Kimian assumption, that emergent moral properties can be causal factors behind other moral properties in a way that involves no downward causation at all.

However, the Kimian assumption in question seems much more plausible when it comes to the judgment-cases: it is plausible to think that moral properties must cause our moral judgments by causing their physical supervenience bases. So, if we accept the judgment-cases as cases of moral causation, then we should also accept that they involve downward moral causation if they also involve emergent moral properties. Even so, though, the derivation of overdetermination from this downward moral causation is dubious because it may not be the case that moral properties would have to be competing causes with their underlying physical properties. The judgment-cases could be genuine counterexamples to the Causal Closure of the Physical Principle needed to generate the causal competition here, which again holds that, for any given physical effect, E, there is some set of physical properties, P, that causes E: perhaps some of our moral judgments, E, result from moral perception and so are caused by moral properties rather than their physical supervenience bases.

Alternatively, there may be no causal competition here because it may not be appropriate to talk about properties as causes rather than as factors implicated in the causation of certain effects. Recall our second causal interpretation of the judgment-cases from earlier, where we understood these cases as those in which moral properties are not, strictly speaking, the causes of our moral judgments even though they *are* directly implicated in the causation of these judgments. The model for this possibility again comes from Ted Honderich’s example of a pear weighing down a scale, which it does in virtue of its weight as opposed to things such as its color. We can understand it as the pear—the object—causing the scale to go down in virtue of certain properties rather than others. These properties are causally relevant factors here—they are implicated in the causal bringing about of the scale going down—but the cause is the object. Or we can understand the cause as an event rather than an object: the event of you putting the pear on the scale causes the scale to go down because of the pear’s weight. Either way we understand it, we again have a case where a property is directly implicated in the causation of something without being the cause. And since the judgment-cases might be ones in which moral properties are directly implicated in the causation of our moral judgments even though they are not the causes of these judgments, moral properties might still have causal powers by being directly implicated in the causation of our judgments even though they are *not* the causes of our judgments and so are *not* in causal competition with other properties. No competition, no overdetermination.

For the sake of argument, however, suppose that downward moral causation does lead to overdetermination. Why buy the assumption that overdetermination is a bad result that warrants rejecting such causation and therefore Emergentist Ethical Naturalism as a whole? In a paper motivated solely by the question of why overdetermination is bad, Theodore Sider locates three possible answers to it that might be able to vindicate the assumption in question.[[50]](#footnote-50) The first is that such overdetermination is metaphysically impossible because the correct theory of causation precludes it. If so, then downward moral causation implies a metaphysical impossibility, which makes such causation—and thus Emergentist Ethical Naturalism as a whole—metaphysically impossible as well, which of course warrants the rejection of both. The second answer is that widespread overdetermination is implausible because it would be a massive coincidence. If so, then downward moral causation implies an implausible phenomenon, which makes such causation—and therefore Emergentist Ethical Naturalism as a whole—implausible as well, which again warrants the rejection of both. And the third answer seems to be another appeal to parsimony: Ockham’s Razor says that we should not believe in explanatory superfluous entities, so since emergent moral properties are overdetermining, explanatorily superfluous entities, Ockham’s Razor implies that we should not believe in them.

Though seemingly plausible, none of these attempts to sink Emergentist Ethical Naturalism succeed. In response to the thought that overdetermination is impossible because the correct theory of causation precludes it, Sider points out that none of the commonly defended theories preclude overdetermination, which is presumably meant to suggest that the correct theory of causation need not actually preclude overdetermination. But if this is true, then there is no longer any basis for maintaining that downward moral causation implies a metaphysical impossibility and thus that such causation—and consequently Emergentist Ethical Naturalism as a whole—is impossible as well. Also, in response to the claim that widespread overdetermination is implausible because it would be a massive coincidence, Sider points out that overdetermination resulting from entities connected by metaphysical necessity would be no coincidence precisely in virtue of such a metaphysical connection. But since this is true, given that the distinct sufficient causes in the context of the current problem, M and P, would be connected by such necessity, the resulting overdetermination would be no coincidence at all. There is thus no longer any basis for maintaining that the overdetermination resulting from downward moral causation would be an implausible coincidence and therefore that such causation—and consequently Emergentist Ethical Naturalism as a whole—is implausible as well.

Lastly there is the argument that, because they would be explanatorily superfluous, overdetermining entities, Ockham’s Razor dictates that we should not believe in emergent moral properties. Yet these grounds for not believing in such properties are insufficient for at least two reasons. First, even if Ockham’s Razor provides *a* good reason to not believe in emergent moral properties by speaking against them, we should not refuse to believe in such properties on account of Ockham’s Razor alone because there may still be other, competing considerations speaking in favor of believing in such properties that outweigh those on the other side. Second, Ockham’s Razor is not a valid epistemic principle in its own right because the mere fact that something would be explanatorily superfluous does not, by itself, constitute a good epistemic reason to not believe in it (e.g., the fact that only one fatal bullet from a firing squad is needed to explain the victim’s death, which renders any other fatal shots explanatorily superfluous, does not give us good reason to not believe that the victim was hit by more than one fatal shot). Ockham’s Razor thus stands in need of vindication, and for this we need good, Razor-independent grounds for thinking that explanatory superfluity indicates probable or actual non-existence, or a lack of sufficient evidence to warrant belief, or something of the sort that clearly constitutes a good epistemic reason to not believe in something.

In the end, I think that these responses available to the emergentist-naturalist demonstrate that the problem of downward moral causation does not, as it stands, spell doom for Emergentist Ethical Naturalism. In order for downward causation to sink the theory, it must clearly have causal overdetermination as a consequence, and such overdetermination must be shown to be sufficiently problematic so as to warrant the rejection of such moral causation and therefore of Emergentist Ethical Naturalism as a whole. So, until these gaps are filled in, the theory should not be rejected in spite of the theoretical benefits that it offers to moral realists.

**7. Conclusion**

In this paper I have reconstructed and defended an emergentist theory of moral properties—Emergentist Ethical Naturalism—that affirms the robust ontological non-reductionism that typically defines ethical “non-naturalism” as well as the naturalist claim that moral properties are natural ones. As the theory purports to demonstrate, the robust ontological non-reductionism that typically defines “non-naturalism” can be part of a coherent, genuinely naturalistic conception of moral properties that stems from a naturalistic worldview that should appeal to both sides of the traditional debate between naturalist and “non-naturalist” moral realists. This theory provides a unifying, emergentist account of several attractive metaethical commitments—those of moral realism, robust ontological non-reductionism, causal efficacy, ethical naturalism, moral non-deducibility, asymmetrical moral dependence, moral determination, and moral supervenience. Particularly noteworthy is its ability to explain, via the emergence of moral properties, why real, objective, sui generis, robustly irreducible moral properties supervene on the lower-level, non-moral properties that ground them as well as how such moral properties are anchored into our naturalistic world. My focus in this paper, however, has been on modestly defending the theory by (1) highlighting the additional theoretical benefits it has for moral realists in virtue of its commitment to causal efficacy, or the claim that moral properties have causal powers, (2) offering support for this commitment in the form of instances of actual moral causation, (3) addressing strong objections to such causation, and (4) contending with the problem of downward moral causation that stems from the theory’s postulation of emergent moral properties as those responsible for the actual moral causation that it relies upon. First, I argued that, by implying that moral properties have causal powers, Ethical Emergentism implies that these properties satisfy an ontological principle demanding that all real things have causal powers and thus does not require the rejection of this principle. I further argued that Ethical Emergentism falsifies the epiphenomenalist premise of an anti-realist syllogism that also utilizes this ontological principle and therefore allows us to reject this premise (and thus the syllogism) without having to construe moral properties as “nothing over and above” their natural, causally powerful grounding properties. Finally, I argued that, by implying that moral properties have causal powers, Ethical Emergentism delivers the causally powerful moral properties that seem to be required for perceptual moral knowledge and thereby gets the possibility of such knowledge off the ground. After discussing these benefits of the commitment to causal efficacy, I offered two kinds of cases of actual moral causation as support for this commitment—what I called “judgment-cases” in which moral properties appear to be causally implicated in our moral judgments along with “compensation-cases” in which moral properties appear to be causally implicated in the bringing about of other moral properties. Then, in order to preserve this support offered for the commitment to causal efficacy, I defended moral causation against five strong objections to it. Lastly, following Jaegwon Kim’s discussion of emergence, I explained how Emergentist Ethical Naturalism faces a problem with downward moral causation and argued that, as it stands, it does not spell doom for the theory. Of course, this theory calls for further probing and a much more comprehensive defense than that given here, but hopefully I have shown that it is an interesting, attractive, and promising one that warrants both.[[51]](#footnote-51)

1. For the classic discussion and critique of analytic identity-naturalism, see G. E. Moore, *Principia Ethica* (New York: Dover Publications Inc., 2004). For a sophisticated version of analytic identity-naturalism that identifies moral properties with disjunctive natural properties that are formed by joining all of the sets of lower-level natural properties that each ground moral ones together with OR-operators as a result of analytically reducing moral concepts to those of these disjunctive natural properties via a network analysis of moral concepts, see Frank Jackson and Philip Pettit, ‘Moral Functionalism and Moral Motivation,’ *The Philosophical Quarterly* 45 (1995), pp. 20-40 and Frank Jackson, *From Metaphysics to Ethics: A Defense of Conceptual Analysis* (New York: Oxford University Press, 1998). [↑](#footnote-ref-1)
2. David O. Brink, *Moral Realism and the Foundations of Ethics* (New York: Cambridge University Press. 1989); Peter Railton, ‘Moral Realism’, *The Philosophical Review* 95 (1986), pp. 163-20; and Nicholas Sturgeon, ‘Moral Explanations’, in G. Sayre-McCord (ed.), *Essays on Moral Realism* (New York: Cornell University Press, 1988), pp. 229-255. This position is also endorsed by Russ Shafer-Landau in his *Moral Realism: A Defense* (New York: Oxford University Press, 2003), yet he considers it to be part of a non-naturalist theory rather than a naturalist one due to epistemological differences between his theory and Brink’s naturalistic theory. [↑](#footnote-ref-2)
3. This example, along with others, can be found in Brink, *Moral Realism and the Foundations of Ethics*, p. 177. [↑](#footnote-ref-3)
4. E. M. Adams, *Ethical Naturalism and the Modern World-view* (Chapel Hill: The University of North Carolina Press, 1960); David Enoch, *Taking Morality Seriously: A Defense of Robust Realism* (New York: Oxford University Press, 2011); Michael Huemer, *Ethical Intuitionism* (New York: Palgrave Macmillan, 2005); G. E. Moore, *Principia Ethica*, Chapter 1; Graham Oddie, *Value, Reality, and Desire* (New York: Oxford University Press, 2005); and Derek Parfit, *On What Matters, Volume II* (New York: Oxford University Press, 2011). As previously noted, Russ Shafer-Landau considers himself to be a non-naturalist realist as well, but even though he endorses the idea that moral properties are sui generis and irreducible, these properties are only irreducible in the sense of not being identical to their natural grounding properties. Under his non-naturalist theory, moral properties are sui generis but not “robustly” irreducible because they are constituted by their natural grounding properties. [↑](#footnote-ref-4)
5. Ryan Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, *Philosophical Studies* 175 (2018), pp. 339-362. [↑](#footnote-ref-5)
6. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 351. [↑](#footnote-ref-6)
7. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, pp. 354-361. [↑](#footnote-ref-7)
8. My defense of this theory, then, is a rather modest one focused on (a) highlighting its additional theoretical benefits in virtue of its commitment to causal efficacy, (b) offering support for this commitment, (c) responding to objections meant to undermine this support, and (d) addressing a troubling problem that the theory faces by postulating causally efficacious, emergent moral properties. A more comprehensive defense of the theory, although certainly called for, is well beyond the scope of this paper. [↑](#footnote-ref-8)
9. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 350. This minimalist conception of metaphysical naturalism comes from Nicholas Sturgeon, ‘Ethical Naturalism,’ in D. Copp (ed.), *The Oxford Handbook of Ethical Theory* (New York: Oxford University Press, 2007), pp. 91-121. Sturgeon arguably conceives of metaphysical naturalism along the same lines in his ‘Moral Explanations’, p. 241. Also, according to Evan Fales, Alvin Plantinga, who is one of the most formidable opponents of metaphysical naturalism, understands it essentially the same way as the view that there are no gods or anything very much like a god. See his ‘Naturalism and Physicalism,’ in M. Martin (ed.), *The Cambridge Companion to Atheism* (New York: Cambridge University Press, 2007), pp. 118-134. The majority of scientific inquiry also seems to presuppose such a minimalist conception of metaphysical naturalism in so far as it seeks only naturalistic explanations of things as opposed to supernatural or spiritual ones (e.g., theistic explanations). At any rate, while this minimalist conception will almost certainly be controversial and invites important questions (e.g., what all count as supernatural or spiritual beings?), it still seems to capture what many people—proponents and opponents alike—have in mind when speaking about “metaphysical naturalism.” [↑](#footnote-ref-9)
10. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 350. This minimal notion of naturalization will also be controversial, and it definitely deserves more treatment than that given here, but there are a few positive things about it that are worth noting before moving on. First, it is metaethically informative and significant. For moral properties that count as natural under this minimal notion are not supernatural or spiritual entities in their own right (e.g., by being properties of God, such as his moral perfection) or dependent on anything, such as God’s will, that is supernatural or spiritual. Metaethical theories that hold moral properties to be natural, then, answer the important metaethical issue of whether moral properties are metaphysically dependent on God’s will by implying that they are not. Second, this minimal notion is inclusive. For what are often regarded as natural properties, such as those that the natural and social sciences study, will count as natural under this minimal notion because they, of course, fit into our naturalistic world. The same is true of any other non-moral, non-evaluative, non-normative, or purely descriptive properties that we know already exist alongside those that the sciences affirm. But even though these properties are correctly counted as natural under this minimal notion of naturalization, they are not necessarily the only ones that can count; instead, any additional properties that can co-exist with these other natural properties and can be coherently combined with them into a continuous naturalistic reality will also count as natural ones. Accordingly, while moral properties can indeed be naturalized by postulating an identity or constitution relation between them and their grounding natural properties, which again renders them “nothing over and above” these natural properties, they can also be naturalized if they are sui generis and robustly irreducible so long as they can be coherently combined with their natural grounding properties into a continuous naturalistic reality. This ultimately means that so-called “non-naturalist” realists can count as ethical naturalists along with identity-naturalists and constitution-naturalists. [↑](#footnote-ref-10)
11. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 347. It is worth noting that by postulating an emergence relation that holds with metaphysical necessity between moral properties and their grounding non-moral properties, this metaethical thesis runs into the objection that emergence relations only hold with nomological necessity (Brian McLaughlin, ‘Emergence and Supervenience,’ in M. Bedau and P. Humphreys (eds.), *Emergence: Contemporary Readings in Philosophy and Science* (Cambridge: MIT Press, 2008), pp. 81-97; Paul Noordhof, ‘Emergent Causation and Property Causation’, in C. MacDonald and G. MacDonald (eds.), *Emergence in Mind* (New York: Oxford University Press, 2010), pp. 69-99). This objection, though, begs the question against Ethical Emergentism. How come properties cannot emerge with metaphysical as opposed to nomological necessity? Even if some emergent properties emerge with only nomological necessity, why are other properties, such as moral ones, precluded from emerging with metaphysical necessity, especially when the relevant bridge laws that connect moral properties to their grounding properties are metaphysically necessary as opposed to contingent? [↑](#footnote-ref-11)
12. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, pp. 347-349. [↑](#footnote-ref-12)
13. Some may understand moral supervenience to be no different than moral determination, but I have chosen to stick with the sloganized version of moral supervenience that, while implied by moral determination, does seem to be a different claim: while moral determination captures important metaphysical relations holding between moral properties and their lower-level grounding properties, moral supervenience rules out the possibility of a certain situation obtaining—namely, one in which multiple objects of proper moral evaluation have different moral properties yet no difference in their lower-level, non-moral ones. [↑](#footnote-ref-13)
14. One or both of these ideas can be found in Mark Bedau and Paul Humphreys, ‘Introduction’, in M. Bedau and P. Humphreys (eds.), *Emergence: Contemporary Readings in Philosophy and Science* (Cambridge: MIT Press, 2008), pp. 1-6; Carl Hempel and Paul Oppenheim, ‘On the Idea of Emergence’, in M. Bedau and P. Humphreys (eds.), *Emergence: Contemporary Readings in Philosophy and Science* (Cambridge: MIT Press, 2008), pp. 61-67; Paul Humphreys, ‘Aspects of Emergence’, *Philosophical Topics* 24 (1996), pp. 53-70; Jaegwon Kim, ‘Making Sense of Emergence’, in M. Bedau and P. Humphreys (eds.), *Emergence: Contemporary Readings in Philosophy and Science* (Cambridge: MIT Press, 2008), pp. 127-152; Jaegwon Kim, ‘Emergence: Core Ideas and Issues’, *Synthese* 151 (2006), pp. 547-559; Sandra Mitchell, ‘Emergence: Logical, Functional, and Dynamical’, *Synthese* 185 (2012), pp. 171-186; Noordhof, ‘Emergent Causation and Property Causation’, pp. 70-71; Timothy O’Connor and Hong Yu Wong, ‘The Metaphysics of Emergence’, *Nous* 39 (2005), pp. 658-678; Timothy O’Connor and Hong Yu Wong, ‘Emergent Properties’, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), <http://plato.stanford.edu/archives/sum2015/entries/properties-emergent/>; Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 343; and Hong Yu Wong, ‘Emergents from Fusion’, *Philosophy of Science* 73 (2006), pp. 345-367. [↑](#footnote-ref-14)
15. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 343. The idea that emergent properties are determined by lower-level ones can be found in Kim, ‘Emergence: Core Ideas and Issues’, pp. 548-550 and Noordhof, ‘Emergent Causation and Property Causation’, pp. 70-71, while the idea that emergent properties are dependent ones can be found in Bedau and Humphreys, ‘Introduction’, p. 1, Kim, ‘Emergence: Core Ideas and Issues’, pp. 550, 556, and Wong, ‘Emergents from Fusion’, p. 346. [↑](#footnote-ref-15)
16. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 343. The idea that emergent properties are irreducible ones can be found in Kim, ‘Making Sense of Emergence’, pp. 129, 132, 141; Kim, ‘Emergence: Core Ideas and Issues’, pp. 548, 554; O’Connor and Wong, ‘The Metaphysics of Emergence’, p. 663; O’Connor and Wong, ‘Emergent Properties’; and Wong, ‘Emergents from Fusion’, p. 346. [↑](#footnote-ref-16)
17. It might be worth stressing that these emergentist claims about consciousness are not theses that I am here asserting to be true. At the same time, I do not reject them as false. The discussion here is non-committal about their truth and only uses them to illustrate what, theoretically speaking, emergent properties are (as I understand them) by describing what would be true about consciousness *if* it were an emergent property. For a full-blown defense of mental emergentism that focuses on the qualitative and intentional features of mental states as emergent phenomena, see O’Connor and Wong, ‘The Metaphysics of Emergence’. For a different defense of emergence that focuses on that which occurs in complex scientific systems, see Mitchell, ‘Emergence: Logical, Functional, and Dynamical’. The success of ethical emergentism, however, does not depend on the success of these other kinds of emergentism. [↑](#footnote-ref-17)
18. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 343. [↑](#footnote-ref-18)
19. Simone Gozzano, ‘Emergence: Laws and Properties: Comments on Noordhof’, in C. MacDonald and G. MacDonald (eds.), *Emergence in Mind* (New York: Oxford University Press, 2010), pp. 100-107; Kim, ‘Making Sense of Emergence’, pp. 129, 141; Kim, ‘Emergence: Core Ideas and Issues’, p. 557; Cynthia MacDonald and Graham MacDonald, ‘Emergence and Downward Causation’, in C. MacDonald and G. MacDonald (eds.), *Emergence in Mind* (New York: Oxford University Press, 2010), pp. 139-168; Mitchell, ‘Emergence: Logical, Functional, and Dynamical’, p. 173; Timothy O’Connor, ‘Emergent Properties’, *American Philosophical Quarterly* 31 (1994), pp. 91-104; Timothy O’Connor and John Ross Churchill, ‘Is Non-reductive Physicalism Viable within a Causal Powers Metaphysic?’, in C. MacDonald and G. MacDonald (eds.), *Emergence in Mind* (New York: Oxford University Press, 2010), pp. 43-60; O’Connor and Wong, ‘The Metaphysics of Emergence’, p. 663; Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 343; and Rex Welshon, ‘Emergence, Supervenience, and Realization’, *Philosophical Studies* 108 (2002), pp. 39-51. [↑](#footnote-ref-19)
20. O’Connor and Wong, ‘The Metaphysics of Emergence’, p. 662 and Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 344. [↑](#footnote-ref-20)
21. Hempel and Oppenheim, ‘On the Idea of Emergence’, p. 62; Kim, ‘Making Sense of Emergence’, pp. 129-132, 141; Kim, ‘Emergence: Core Ideas and Issues’, p. 551; Mitchell, ‘Emergence: Logical, Functional, and Dynamical’, p. 173; and Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 344. [↑](#footnote-ref-21)
22. Kim, ‘Making Sense of Emergence’, pp. 129-132, 141; Kim, ‘Emergence: Core Ideas and Issues’, p. 551; and Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 344. [↑](#footnote-ref-22)
23. Kim, ‘Making Sense of Emergence’, pp. 129-132; Kim, ‘Emergence: Core Ideas and Issues’, p. 551; and Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 344. [↑](#footnote-ref-23)
24. C. D. Broad, *The Mind and its Place in Nature* (New York: Harcourt, Brace and Company, Inc, 1925); Kim, ‘Emergence: Core Ideas and Issues’, p. 552; and Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, pp. 344-345. [↑](#footnote-ref-24)
25. Stringer, ‘Realist Ethical Naturalism for Ethical Non-Naturalists’, p. 354. [↑](#footnote-ref-25)
26. For more on the importance that perceptual moral knowledge may have for ethical theorizing, see Sarah McGrath, ‘Moral Perception and Its Rivals’, in A. Bergqvist and R. Cohen (eds.), *Evaluative Perception* (New York: Oxford University Press, 2018), pp. 161-182. [↑](#footnote-ref-26)
27. For accounts of moral perception that reject this plausible assumption and argue instead that perception of moral properties—and thus perceptual moral knowledge—can happen so long as moral properties are adequately connected with other properties, such the lower-level natural properties that ground them, that do stand in a direct causal relation to our epistemic equipment, see Robert Audi, ‘Moral Perception and Moral Knowledge’, *Proceedings of the Aristotelian Society Supplementary Volume* 84 (2010), pp. 79-97; Andrew Cullison, ‘Moral Perception’, *European Journal of Philosophy* 18 (2010), pp. 159-175; and Justin McBrayer, ‘Moral Perception and the Causal Objection’, *Ratio* 23 (2010), pp. 291-307. On these intriguing accounts, it is possible to perceive causally impotent moral properties, and so there can be perceptual moral knowledge even if moral properties cannot causally affect us. [↑](#footnote-ref-27)
28. It is important to note, contra to J. L. Mackie’s position in *Ethics: Inventing Right and Wrong* (New York: Penguin Books, 1977), that realists need not be committed to the idea that the requisite epistemic equipment for such knowledge must consist in or include a weird faculty of moral perception (Brink, *Moral Realism and the Foundations of Ethics*, p. 110; Sarah McGrath, ‘Moral Knowledge by Perception’, *Philosophical Perspectives* 18 (2004), pp. 209-228). Furthermore, the idea that moral properties can causally affect our epistemic equipment does not require a commitment to the existence of special moral particles that make this possible (Brad Majors, ‘Moral Explanation and the Special Sciences’, *Philosophical Studies* 113 (2003), pp. 121-152). [↑](#footnote-ref-28)
29. For a defense of moral causation that argues for genuine, irreducible causal laws in which moral properties figure, see Majors, ‘Moral Explanation and the Special Sciences’. [↑](#footnote-ref-29)
30. Brink, *Moral Realism and the Foundations of Ethics*, p. 186; Brad Majors, ‘Moral Explanation’, *Philosophy Compass* 2/1 (2007), pp. 1-15; and Sturgeon, ‘Moral Explanations’, p. 234. [↑](#footnote-ref-30)
31. Sturgeon, ‘Moral Explanations’, p. 234. [↑](#footnote-ref-31)
32. Of course, one might object to these interpretations on the grounds that the “because” in Sturgeon’s answer is not a causal one. But if so, what kind of “because” is it? At this point the onus is on the imagined critic here to specify what kind of “because” we have here and why it is this kind rather than a causal one. [↑](#footnote-ref-32)
33. Ted Honderich, ‘The Argument for Anomalous Monism’, *Analysis* 42 (1982), pp. 59-64. [↑](#footnote-ref-33)
34. Brink, *Moral Realism and the Foundations of Ethics*, p. 187; Majors, ‘Moral Explanation’, p. 3; Railton, ‘Moral Realism’, pp. 191-192; and Sturgeon, ‘Moral Explanations’, p. 245. [↑](#footnote-ref-34)
35. The need for a symmetrical explanation of the behavior here was suggested by Dallas Amico. [↑](#footnote-ref-35)
36. See Railton, ‘Moral Realism’, pp. 191-192, however, for a case where the causation between the moral properties of society and the social unrest need not be mediated by the beliefs of the restless. If this case pans out, then we would indeed have a fresh supporting example of moral causation. [↑](#footnote-ref-36)
37. Sturgeon, ‘Moral Explanations’, pp. 243-244. [↑](#footnote-ref-37)
38. Terence Cuneo, ‘Moral Facts as Configuring Causes’, *Pacific Philosophical Quarterly* 87 (2006), pp. 141-162. [↑](#footnote-ref-38)
39. My rejection of Cuneo’s case of moral causation rests on the intuitively attractive idea that we can smoothly separate non-moral qualities from moral ones, which Cuneo might claim is too dubious to rely upon. If this turns out to be true and Cuneo’s case of moral causation turns out to be genuine, then so much the better for the commitments to moral causation and causal efficacy! [↑](#footnote-ref-39)
40. For this to hold up, it must be the case that Walter can pay the relevant sum of money, since he cannot be obligated to do what he cannot do. If we then assume that he cannot pay the money, then we would have to modify the example by making the compensation that he owes to his victim something that he can pay. [↑](#footnote-ref-40)
41. Although he does not explicitly formulate this objection, we could attribute it to Gilbert Harman, ‘Ethics and Observation’, in G. Sayre-McCord (ed.), *Essays on Moral Realism* (New York: Cornell University Press, 1988), pp. 119-124. [↑](#footnote-ref-41)
42. Majors (2003) mentions this objection in the context of defending moral causation. He also mentions an overdetermination objection to moral causation that I will effectively deal with later on when addressing the problem of downward causation. And once again, though Harman does not explicitly formulate this objection, he may have something like it in mind when he points out the apparent lack of a possible mechanism that would allow moral properties to have causal effects on our perceptual apparatus. [↑](#footnote-ref-42)
43. Cuneo, ‘Moral Facts as Configuring Causes’, p. 142; Majors, ‘Moral Explanation and the Special Sciences’, p. 137. [↑](#footnote-ref-43)
44. Once again, although Harman does not explicitly formulate such a complaint, we could plausibly attribute it to him. [↑](#footnote-ref-44)
45. It is worth noting here that realists could accept something very much along the lines of this “anti-realist” explanation of our intuition about the hoodlums and reject the “realist” explanation that cites a moral property as a causal factor behind this intuition. Realist-epiphenomenalists that reject causally powerful moral properties, and that believe instead that such properties are causally impotent yet nevertheless real, will presumably accept something like this “anti-realist” explanation and reject the “realist” one, and so the labels here may be somewhat misleading. They are, however, useful in the present context to differentiate between explanations that make no mention of moral properties, which are obviously going to be favored by anti-realists, and those that do cite such properties, which of course will only appeal to realists. [↑](#footnote-ref-45)
46. Brian Leiter, ‘Moral Facts and Best Explanations’, *Social Philosophy and Policy* 18 (2001), pp. 79-101. [↑](#footnote-ref-46)
47. It is assumed here that the physical level is ontologically fundamental, but one could understand this as epistemically fundamental instead without doing any violence to the argument. [↑](#footnote-ref-47)
48. Kim, ‘Making Sense of Emergence’, pp. 141-143, 149-150 and Kim, ‘Emergence: Core Ideas and Issues’, pp. 557-558. [↑](#footnote-ref-48)
49. It is important to note that the Causal Closure of the Physical Principle used to generate this bad consequence is not the same as the Causal Closure of the *Natural*, which would instead maintain that every effect has some sufficient naturalistic cause. This latter principle will clearly be true under Emergentist Ethical Naturalism since the world is a naturalistic one with only natural stuff, including causation and its relata: since only natural things exist, only such things can have effects, and so every actual effect must have a sufficient naturalistic cause. This alternative closure principle does not imply the Causal Closure of the Physical because it does not require that naturalistic causes be physical ones, or ones at the lowest level of physics. Of these two closure principles, it is the Causal Closure of the Physical that is particularly needed for this objection to go through because only this one guarantees that causally efficacious, emergent moral properties have causal competitors at the fundamental physical level. [↑](#footnote-ref-49)
50. Theodore Sider, ‘What’s So Bad about Overdetermination?’, *Philosophy and Phenomenological Research* 67 (2003), pp. 719-726. [↑](#footnote-ref-50)
51. I would like to thank Craig Agule, Dallas Amico, David Brink, Cory Davia, an audience from the Moral and Political Philosophy Seminar at UC San Diego, an anonymous referee for another journal, and two anonymous reviewers for this journal for valuable feedback on earlier versions of this paper. I owe a very special thanks to Matthew Fulkerson and to the associate editor of this journal, who each provided valuable feedback on two earlier versions of the paper. [↑](#footnote-ref-51)