

DIVINE MOTIVATION THEORY

TERENCE CUNEO
Calvin College

Theistic views in the foundations of ethics receive relatively little attention these days. While there are probably numerous reasons for this, the relative neglect of these positions is, at first blush, somewhat surprising. After all, on the face of it, theistic views in ethics promise considerable advantages vis-à-vis their non-theistic counterparts. For one thing, theistic views appear capable of capturing handily the robustly realist appearances of moral thought and practice, including the categorical bindingness of moral obligation. For another, such views appear to be able to explain why we have reason to care about being moral, as they closely connect—to use Kant's terminology—virtue and happiness. Given these putative advantages, one wonders: why the relative neglect?

A plausible answer is that many philosophers believe that the options for theistic ethical theories have been both exhausted and found wanting. The relevant options are usually presented as such: one can be either a theistic voluntarist or a theistic Platonist. The problem with voluntarist views are twofold. In the first place, paradigmatic instances of this position suffer from a lack of generality inasmuch as they offer a theistic account not of the moral domain in general, but only of the nature of moral obligation. Moreover, voluntarist views threaten to forfeit important aspects of the objectivity of moral obligation. For, according to paradigmatic examples of the position, the wrongness of an act-type such as *being a wanton murder* is only contingently displayed by such an act-type—it being possible for God not to have forbidden the performance of an act-type of this kind. The problem with theistic Platonist positions, by contrast, is that they preserve the objectivity of morality only by making it difficult to see why God matters for morality. If there is an objective moral domain that exists independently of the will or nature of God, it seems that God's role is at most to convey moral truths to us or to ensure that human agents get their just deserts.

Fundamental to Linda Zagzebski's *Divine Motivation Theory* is the contention that the options for theistic ethical theories have not been exhausted.¹ In Zagzebski's view, there is a virtue-based theistic view—to this point almost entirely overlooked by philosophers—that deserves the attention of non-theists and theists alike. Non-theists should take note of the view's explanatory power; for, if correct, it promises to deliver a great deal of what most ethical theories have endeavoured to capture. Theists should find the position congenial for several reasons. If correct, the view offers a general account of the moral domain that preserves both the objectivity of morality and makes God

1. *Divine Motivation Theory*, by Linda Trinkaus Zagzebski (Cambridge University Press, 2004. xviii + 410 pp).

essential to morality. Moreover, the position promises to make significant headway on thorny problems that have plagued theistic positions, such as the problem of evil.

Zagzebski presents her view in three stages, which correspond to the three parts of *Divine Motivation Theory*. Part one, 'Motivation-based virtue ethics', develops a general type of virtue theory according to which good motivations are (conceptually and ontologically) basic. Part two, 'Divine Motivation Theory', develops a theistic version of this theory, illustrating some of its applications to several problems in the philosophy of religion. Part three, 'Ethical Pluralism', addresses the issue of how a motivation-based view can handle the phenomenon of deep and protracted ethical conflict. These sections unfold as follows.

Since the publication of John Rawls's *A Theory of Justice*, moral philosophers have been more or less consumed with the task of constructing an ethical theory that makes one or another evaluative notion or property primary. For some time, philosophers have assumed that there are three options from which to choose: one can be a deontologist and thus accept the notion of the right as primary. Or one can be a teleologist and accept the notion of the good as primary. Or, finally, one can be a virtue theorist and hold that the notion of a virtue is basic. Zagzebski opts for the last option, albeit with a twist. (Zagzebski does not entertain the possibility of accepting a 'no-conceptual-priority' view, according to which there is no evaluative concept that is most basic.) Her claim is that the concept of a virtue itself should be explicated in terms of the concept of a good motivation. As indicated, this 'motivation-based' virtue theory has two components.

First, the position gives us an account of the dependence relationships between evaluative concepts. As Zagzebski develops the view, we begin by supposing that a good motive is conceptually basic. We then define a virtue as a character trait constituted by a good motive disposition that reliably brings about those states of affairs that are the ends of that motive. Finally, we define other evaluative concepts such as a virtuous act, a right act, a wrong act, a duty, what is good for an agent, and a flourishing life in terms of the motivations of the virtuous agent. Roughly put, a virtuous act is one that expresses a virtue; a right (permissible) act is an act that a virtuous agent might do; a wrong act is one that a virtuous agent would not do and would feel guilty about having performed had he performed it; a duty is an act a virtuous agent characteristically would do and would feel guilty about having not performed, had he not performed it. Thus understood, a distinctive feature of motivation-based virtue ethics is that it forgoes the traditional attempt to define the virtues as those traits that are conducive to or constitutive of eudaimonia or human flourishing. Rather, taking a page from Aristotelians such as John McDowell, Zagzebski proposes that we think of the relationship between the virtues and human wellbeing in the opposite fashion: human flourishing is simply what a virtuous agent would say it is. This move is bound to raise questions, but it does have some apparent advantages, among which is not postulating a broadly Aristotelian 'metaphysical biology' according to which human nature has a *telos* or an end at which we naturally aim.

The second feature of motivation-based virtue ethics is that it offers an account of the relationships of metaphysical dependence between evaluative features. In this case, the metaphysical dependence relations mirror the conceptual dependence relations. Accordingly, properties such as *being a virtuous agent*, *being a virtuous action*, *being a right action*, and so forth supervene on the property of *being a good motive*. A consequence of this view is that the end of an action—typically, a state of affairs that an action aims to bring about—is not what makes it the case that that action is morally good or bad. Rather, what makes it the case that the end of an action is morally good is the reason why it is aimed at. And the reason why an end is aimed at, according to Zagzebski, is the motive that aims to bring it about—where a motive is not a state of affairs that an action aims to bring about, but a mental state that an action expresses. As such, motives are reasons in a two-fold sense: they both cause actions and (when all goes well) justify those actions that they cause. For example, according to this view, when asked: ‘Why did you give up your summer vacation to aid your ailing mother?’ the answer will appeal to a motive: ‘Because doing so expresses love for my mother.’ This answer is supposed to both explain and justify performing the action in question.

As it stands, this type of view places a heavy explanatory burden on the notion of being a good motive. What exactly is this mental state that explains and justifies action? Zagzebski devotes the book’s second chapter to this topic. The view she defends is that motives are emotions. Emotions, in turn, are affective states with intentional objects. More exactly, according to Zagzebski, when an agent is in the state of having an emotion, she sees the intentional object of that emotion as falling under what she calls a “thick affective concept”—a concept such as ‘being lovable’, ‘being pitiful’, ‘being dangerous’, ‘being contemptible’, and so on. Thus described, emotions are not, as some philosophers have believed, brute affective impulses outside the realm of evaluative assessment. Rather, emotions can be good, Zagzebski claims, and are so when they “fit” their intentional object. And an agent’s emotion fits an intentional object just in case that object has the thick affective property that that agent sees it as having in that emotional state. Indeed, Zagzebski’s claim is that when an emotion thus fits its object it is not only good, but also intrinsically good, for its goodness is not derived from any other good thing.

All of this raises a natural worry. Won’t the motivation-based view thus described be hopelessly circular? After all, it is difficult to see how we could explicate the concept of a good motive without making essential reference to a virtuous agent, for a good motive is arguably best understood as one that would move a virtuous agent to act. But, according to a motivation-based view, the order of explication is supposed to go the other way around: virtuous agents are supposed to be defined as those that have good motives. Zagzebski’s strategy for addressing this concern is to borrow from theories of direct reference as they’ve been developed by thinkers such as Kripke and Burge. According to the view Zagzebski favours, we distinguish between two types of definitions of evaluative concepts.

According to so-called lexical definitions, we begin by fixing the reference of an expression such as ‘good person’ without the attributive use of any

concepts, but simply with a demonstrative thought. The good person is the person *like that*. A good emotion, likewise, is an emotion a person *like that* would have in a given situation. A virtue is a disposition to have emotions *like that* and to be motivated to bring about such ends that good persons with such emotions intentionally and reliably bring about, and so on. If Zagzebski is right, by demonstratively identifying moral exemplars in this way, we both avoid a circle of conceptual definitions and fix the meaning of expressions such as ‘good person’. Having done this, we can offer so-called substantive definitions of fundamental evaluative concepts. According to these definitions, a good emotion is an emotion that fits its intentional object, good persons are good (in part) because they have good emotions, a good outcome is the end of a good motive, and so on. Zagzebski favours this way of proceeding because she maintains that our substantive definitions in principle could be mistaken. In principle, for example, our substantive definition of water as H₂O could be wrong (perhaps there is a deeper causal structure to this stuff that we’ve yet to discover). So also the motivation-based virtue theorist’s substantive definition of a good agent could be wrong. The safe bet, then, is to take the indexical definitions as primary and build from these.

In summary, the book’s first section is dedicated primarily to the project of tracing the relations of dependence among our evaluative concepts. The book’s second part is primarily concerned with the project of developing a particular metaphysic of value and illustrating its philosophical payoff. The central claim of this section is that the most compelling version of motivation-based virtue ethics is one in which it is God’s motives that determine the value of objects of moral assessment.

Suppose we assume that God is as traditional theists claim: a person who displays a variety of great-making properties such as being omnipotent, omniscient, impeccable (being unable to sin), everlasting, and so on. If God has these features, then God acts. And if God acts, then God has motives. Zagzebski suggests that God’s motives are best understood to be emotions or at least emotion-like states of various kinds. So, consider God’s having created the world. What could have motivated God to do this?

According to Zagzebski, God did not create this world because God scanned the possibilities and saw that this world would be good were God to create it. Nor did God create this world because God aimed to satisfy one of God’s desires. Rather, Zagzebski suggests that “God’s motive to create was what it was simply because God is God and because as God He possesses certain motives . . . that impel Him to act” (p. 217). So, creation was not done for the sake of anything; it was simply an expression of God’s love, a quality that God expresses because it is God’s nature to do so. As Zagzebski puts it in one place, God’s actions are like those of an artist who creates works of beauty, but who cannot be said to create these works because they would be beautiful (p. 313).

The position that emerges is one in which God’s nature is metaphysically prior to morality. Somewhat more exactly, Zagzebski maintains that a certain range of God’s motivations are intrinsic features of God’s nature. And what is good as an end or an outcome is good because God is (or could be)

motivated to bring it about. If Zagzebski is right, this position has numerous advantages over its rivals such as the Divine Command theory, among which is the fact that it preserves the objectivity of morality. Since certain motives are essential to God, it cannot be the case that, say, an act-type such as *being a wanton murderer* is not morally bad. And, important for Zagzebski's purposes, this view appears to highlight the importance of the Christian doctrine of the incarnation. For fundamental to a motivation-based virtue ethics is the idea that our task as moral agents is to imitate moral exemplars. If God is the supreme exemplar and has become incarnate in the person of Christ, then there is an agent whom we can imitate and who is also the supreme moral exemplar.

Arguably, the most interesting section of *Divine Motivation Theory* is the application of this view to puzzles in philosophical theology. There are three puzzles with which Zagzebski is concerned. The first is how to reconcile God's omnipotence with God's impeccability. In brief, the worry here is that if omnipotence implies having maximal power, then it appears that an impeccable being cannot be omnipotent, since it lacks the power to sin. The second puzzle is related to the first. If God is impeccable, then God appears to lack morally significant freedom. But if God lacks this sort of freedom, it is difficult to see in what sense God is morally praiseworthy, for God cannot but be morally good. Finally, there is the issue of how to reconcile God's impeccability, omnipotence, and omniscience with the amount and types of evil in the world.

To solve the first two puzzles, Zagzebski makes the following conceptual manoeuvre: maintain that God's great-making features are perfections of God, but deny that, if a trait is a perfection, then its possessor must have it to the highest degree. Better, suggests Zagzebski, to hold that God's great-making properties are perfect ways of having properties of certain kinds. If Zagzebski is right, the perfection of power is power in the degree, manner, and scope that a perfectly good God is motivated to have. Likewise, the perfection of freedom is having freedom in the degree, manner, and scope that God is motivated to have. Understood thus, the first two puzzles disappear. For only if we are thinking of perfect power as maximal power and perfect freedom as including the power to do evil do these puzzles seem threatening.

Zagzebski's response to the problem of evil takes a slightly different turn. One formulation of the problem of evil runs thus:

- (1) A perfectly good being would be motivated to prevent every evil.
- (2) An omnipotent being would be able to prevent every evil.
- (3) So if there were a being who was perfectly good and omnipotent, this being would be motivated to prevent every evil and would be able to do so.
- (4) If there were a being who was motivated to prevent every evil and was able to do so, evil would not exist.
- (5) But evil does exist.
- (6) Therefore, there is no being who is perfectly good and omnipotent. The Christian God does not exist.

However, Zagzebski maintains that premise (1) of this argument can be read in two ways. The argument assumes that it should be read as:

- (1a) A perfectly good being is one that is motivated to prevent whatever is evil.

But if Divine Motivation theory is true, (1a) is false—at least if this premise is read to state that God is motivated to prevent evil because it is evil. The following reading of (1), however, is true according to Divine Motivation theory:

- (1b) Evil is whatever a perfectly good being is motivated to prevent.

When we modify each of the argument's premises to accommodate this reading, it turns out that premises (2)—(4) are also true. Premise (5), however, becomes

- (5b) Things exist that an omnipotent and perfectly good being would be motivated to prevent.

But, says Zagzebski, we have no reason to think this claim is true. In fact, we have reason to believe it is false, for there is nothing belonging to the category of outcomes such that God is motivated that they not exist.

There is, I believe, a problem here. Presumably, (1b) should be read as a biconditional, for only if it is read thus does (5b) capture the sense of (5). But, if (1b) is read as a biconditional and (5b) is false, it follows that evil does not exist, which is an implausible result that Zagzebski repudiates (p. 312). If, however, (1b) is read as the conditional

- (1b*) If a perfectly good being is motivated to prevent something, then it is evil,

then it is not (5b), but premise (1) of the original argument that should be rejected. However, if (1) is rejected, and there are evil things, it follows that there are evil things that God is not motivated to prevent. While this may be true, it is now difficult to see how Divine Motivation theory offers us a *solution* to the problem of evil. It appears to shed no light on *why* there are evils that God is not motivated to prevent.

Suppose, however, that Divine Motivation theory provides the types of theoretical benefits that Zagzebski claims. There remains a worry, and that is how the view can handle moral conflict. For in its generic form, motivation-based virtue theory enjoins us to imitate moral exemplars, some of whose views and behaviour are not consistent with one another. In the third part of her book, which consists of only a single chapter, Zagzebski suggests that the key to addressing this issue is to distinguish different perspectives we can take with respect to the self. Appealing to the first-person perspective to solve disagreement, Zagzebski claims, is not promising, as the judgements made from this perspective are frequently unreliable: we are often unable to see

beyond our own failings and epistemic blind spots. Appealing to the third-person perspective, however, is not ideal either, as this perspective is often too removed from our actual concerns to be helpful. Without explicitly saying as much, Zagzebski's own proposal is a position that mirrors certain features of Aristotle's account of friendship: when attempting to solve serious ethical disagreement, appeal to the second-person perspective, which "involves an encounter between two persons' first-person perspectives where each of them understands and appreciates the other" (p. 375).

The foregoing, at any rate, indicates the broad lines of argument in Zagzebski's book. When considered as a unified whole, *Divine Motivation Theory* may have its flaws. But whatever else might be said about it, Zagzebski's book cannot fairly be accused of being *dull*. Teeming with ideas, innovative lines of inquiry, and interesting dialectical moves, this is an extraordinarily creative book. And yet there is so much of interest going on in the book—ranging from treatments of the nature of emotions, intrinsic goodness, the nature of personhood, the problem of evil, the value of narrative, to an exploration of the problems of religious diversity—that it can be difficult at times to see how the discussion hangs together and why the book unfolds as it does.

To illustrate: at the outset of her book, Zagzebski writes that there are three sets of puzzles that drive her project. The first lies in the area of philosophical psychology. It asks: How could it be that moral judgements both express moral propositions and are tied by conceptual necessity to appropriate motivation? The second puzzle lies in the metaphysics of value and raises the following question: How can we explain the difference between descriptive and evaluative properties? The third set of puzzles is comprised of the conundrums in philosophical theology canvassed earlier. These puzzles query: How can we reconcile God's impeccability with God's omnipotence, God's freedom with God's impeccability, and God's goodness with the presence of evil?

It is difficult to see that these puzzles genuinely drive Zagzebski's discussion. The treatment of the first puzzle, for example, occupies seven pages of the book, and is tucked away in a chapter entitled 'Acts and Obligation'. And after having been introduced in Chapter 1, the second puzzle is hardly explicitly mentioned in the remainder of the book, receiving attention only on two pages in a conclusion to the book's first part. In fact, I found Zagzebski's treatment of the second puzzle doubly perplexing. In addition to having difficulty grasping the philosophical import of the puzzle—what exactly hangs on the issue of being able to distinguish precisely the descriptive from the evaluative domains?—I could not see that Zagzebski's own view marks the relevant distinction. Nor do I see a way to extend the motivation-based view so that it would serve to mark the distinction. Were one to extrapolate from what Zagzebski says, one might think that, according to the motivation-based view, an evaluative feature is either an object that could fit an emotion or an entity whose existence could be explained by the fact that an object fits an emotion. This account, however, would be overly capacious, since all sorts of non-evaluative features can be explained by the fact that an object fits an emotion—an agent's feeling delight being among them. Feeling delight, however, is not itself plausibly viewed as being an evaluative feature.

That said about the scope and structure of the book, I want to close by voicing some concerns about the type of position that Zagzebski defends.

Definitive of divine motivation theory is the thesis that good motives are explanatorily basic, for we explain the value of other moral features such as *being a right act*, *being a good state of affairs*, and so forth in terms of such motives. "All moral properties of persons, acts, and the outcomes of acts," says Zagzebski, "are derived from good emotions" (p. 178). I have two worries about what Zagzebski says on this score, both of which concern the explanatory role of thick moral features.

Here is the first worry: when explicating what a good motive is, Zagzebski utilizes the notion of a thick moral feature. Good motives are what correspond to such features. But in the subsequent discussion, nothing is said about how we should account for the presence of such features. Indeed, in the paragraph after the quotation just cited, Zagzebski writes: "I have not said that the value of emotion is prior to the value of states of affairs. I have purposely refrained from saying where the value of the thick properties that we perceive in our emotions come from. That issue will be addressed in Part III" (p. 178). (Zagzebski speaks of how to account for the 'value' of thick moral properties. I take this to be another way of speaking about how to account for the fact that things have this type of 'thick' value; cf. p. 5.) Now if the outcomes of acts are states of affairs, as Zagzebski claims (p. 18), it is difficult to see how to reconcile what she says here with what she says elsewhere. But let that pass. What I am interested in stressing is that when we consider a wide range of thick moral properties such as *being merciful*, *being courageous*, *being wicked*, *being evil*, and so on, it's evident that these properties are response-involving in the following sense: agents do not display these properties merely because some agent *takes* them to have these properties. Such properties are not *conferred* on agents. Rather, agents display these properties because they respond (or are reliably disposed to respond) to reasons in various ways. Agents are merciful, for example, inasmuch as they respond to the suffering of others in certain appropriate fashions. Agents are evil (in part) because they do not so respond. Call moral properties of this sort 'response-involving' thick moral properties.

Arguably, though, this account of thick moral properties makes trouble for the claim that good motives are explanatorily or metaphysically basic. For consider how a chain of explanatory relations would go. We begin by explaining a good motive in terms of a thick moral property: a motive is good insofar as it fits some intentional object that displays a thick moral property. If a wide range of thick moral properties are response-involving, then we explain what it is for something to have such a property in terms of reasons: an agent, for example, displays a response-involving thick moral property inasmuch as she responds in an appropriate or inappropriate way to reasons of a certain range. But if motivation-based theories are correct, we explain reasons such as obligations and permissions in terms of good motives: something is an obligation, for example, inasmuch as it is something an agent with a good motive would do. However, if this is so, we've explained good motives in terms of reasons and vice-versa; there is no evident sense in which either type of entity is

explanatorily more basic than the other. Admittedly, explanatory circularity of this sort doesn't preclude our defining evaluative features in terms of good motives. Zagzebski doesn't hold that good definitions must mirror the explanatory relations in moral reality. Nor does it preclude a view according to which there is no explanatory priority among moral features—such features being at least partially explained in terms of one another. But it does rule out the view that Zagzebski appears to want to defend according to which the ontological order mirrors the conceptual one, such that good motives explain the presence of other moral features and not vice-versa (cf. pp. 5–6, 105).

Still, the suggestion in the passage just quoted is that in part III of *Divine Motivation Theory* we'll receive an explanation of what accounts for the presence of thick moral features. And, indeed, Zagzebski indicates that by bringing in God's affective states, we can explain where the value of thick moral properties comes from. "All value outside of God," writes Zagzebski, "derives from the emotions of God" (p. 386). According to Zagzebski, the "thick properties that the world has are those perceived by God. Their existence depends upon God's affective perception. The loveability of the world comes into existence with God's loving it . . ." (p. 224).

One wonders whether bringing God into the picture in this way helps, however. In the first place, it seems wrong to claim that God confers response-involving thick moral properties on agents. Agents are not merciful, for instance, simply because God takes them to be. Rather, they're merciful because they respond to reasons of a certain range in certain ways. And even if one acknowledges that there are thick evaluative properties such as *being loveable* that are not response-involving, it is unclear what exactly Zagzebski's view is regarding God's relation to the instantiation of these properties. After having acknowledged that the view she defends bears a close resemblance to that of the Lutheran theologian Anders Nygren, Zagzebski pulls back from a natural interpretation of what she says:

I think that Nygren believes that God creates a contingent world and then loves it, and the loving endows it with value. In my view, that is at best misleading. The property of loveability is the intentional object of the divine emotion of love; it is not caused by the emotion any more than the emotion is caused by loveability. (p. 225)

Assuming that Zagzebski is using the term 'cause' in a general way to speak of a determination relation, this is perplexing.

And this leads me to a second worry: I do not understand how Zagzebski is thinking about the way in which thick evaluative properties such as *being loveable* fit into the framework of a divine motivation theory.

On the one hand, Zagzebski repeatedly appears to claim that all value emanates from God: "the outcome of the Creation . . . is good because it is the end of God's motive" (p. 217), the metaphysical version of motivation-based theories claiming that "a perfectly good deity can confer value on the world" (p. 386). And, yet, in the passage cited just earlier, Zagzebski seems to deny that God's loving something confers value on it. On the other hand, at almost

every turn, Zagzebski eschews a Platonist view according to which God does not confer such properties on things, but recognizes them. The divine conferral view and Platonism appear, however, to exhaust the available options. And, as far as I can see, Zagzebski unequivocally accepts neither of these views. (Zagzebski shows no sympathy with a view according to which moral properties are to God as the secondary-quality theorist believes colours are to us—powers that elicit God's emotional responses in certain ways.)

There seem to me, then, two problems regarding thick moral features for motivation-based virtue ethics—the metaphysical version of the position fails to offer an explanatorily informative account of their nature and the theistic version of the view offers no positive account of God's relation to such properties.

I do not think these problems are insurmountable. But it seems to me the divine motivation theorist has to make a choice. One option is to accept the official view that good motives are explanatorily basic. But if she does so, then the divine motivation theorist has not offered us a position regarding what accounts for the fact that things display thick moral features. The second option is to accept the view that God confers properties such as *being loveable* on creatures by virtue of loving them. If the divine motivation theorist takes this route, though, then good motivational states are not explanatorily basic, since it is false that something displays a thick moral feature by virtue of the fact that there is some intentional object that fits God's emotional states. Furthermore, if we agree that persons have value by virtue of the fact that God loves them, it is difficult to see why we should accept the basic structure of the motivation-based package. That persons have value to a high degree seems *itself* sufficient to ground reasons and obligations to them. There is no evident need to account for the existence of such things by appealing to motives that fit their intentional objects.²

2. Thanks to Michael Schwiigger for helpful conversation regarding this review.