

1. Introduction

As rational agents, we are subject to the demands of normative reasons – considerations that count in favor of (or against) actions we might perform, as well as beliefs, intentions, and other attitudes we might hold. Furthermore, for us to be good agents, it isn't enough that our actions and attitudes merely *accord* with what our reasons require. Rather, we must actually (and correctly) *respond* to those reasons in acting, believing, feeling (etc.) as we do.

Correctly responding to normative reasons is a phenomenon of utmost normative importance with regard to a variety of achievements. In this paper, I will be particularly concerned with the achievement of *rationality*.¹ According to an increasingly influential family of views, rationality is a matter of correctly responding to one's reasons.² On such views, for any potentially rational action or attitude, there is a set of normative reasons – call them rationalizing reasons – such that when our actions and attitudes correctly respond to those reasons, that action or attitude is thereby rational.

Call such theories reasons-responsiveness theories of rationality. I think some version of a reasons-responsiveness theory is true. However, there's a serious problem with many extant versions: they rely on accounts of correctly responding to reasons that cannot explain a central feature of rationality. In acting or holding attitudes rationally, agents are *creditworthy* for successfully acting or holding attitudes in accordance with their reasons. If rationality consists in correctly responding to reasons, then correct responsiveness to reasons must be able to explain this central feature of rationality.

My goal in this paper is to develop an account of correctly responding to reasons that satisfies this constraint, and in doing so, to make progress toward the right version of the reasons-responsiveness theory. The paper proceeds as follows. I begin in §2 by providing broad characterizations of rationality and correctly responding to reasons. In §3, I consider an account of correctly responding to reasons

¹ Philosophers have also offered reasons-responsiveness accounts of other achievements, such as morally worthy action. See, for example, Arpaly (2000) and Markovits (2010). In other work, I defend similar accounts of both moral worth (Singh, forthcoming) and knowledge (Singh, ms).

² See, for example, Kiesewetter (2017), Kauppinen (forthcoming), and Lord (2018).

according to which it is a matter of manifesting reasons-sensitive dispositions. In §4, I argue that if rationality consists in correctly responding to reasons, this dispositional account cannot succeed. In §5, I sketch my own account of correctly responding to reasons, according to which correctly responding to reasons requires representing those reasons as the reasons they are. Finally, I conclude in §6 by drawing out two significant implications of my account for the theory of rationality.

2. Preliminaries

Before getting into the more substantial discussion that will occupy the rest of the paper, it will be helpful to provide broad characterizations of its subject matter. My goal in this section is to provide characterizations of both rationality and correctly responding to reasons that can be agreed upon by everyone who is sympathetic to the reasons-responsiveness approach.

2.1. Rationality

The term ‘rationality’ gets used in different ways. Sometimes, it’s used to refer to a property that sets of actions and/or attitudes have when they hang together in the right way. This sense of rationality is usually called *structural* rationality, or *coherence*. This is not the sense of rationality that reasons-responsiveness theories of rationality are in the first instance concerned with. Reasons-responsiveness theories are in the first instance concerned with what is often called *substantive* rationality.³

Unlike structural rationality, substantive rationality isn’t characterized as a property that a set of actions and/or attitudes has when it meets certain structural criteria (criteria regarding the relations between its members). Instead, it’s characterized as a property that a single action or attitude has when it meets certain substantive criteria – that is, criteria about what particular actions or attitudes are called for given the agent’s normative situation.

While reasons-responsiveness theorists are in the first instance concerned with substantive rationality, they’re not necessarily silent about structural rationality. For example, Benjamin Kiesewetter (2017)

³ The terminology of structural and substantive rationality has its roots in Scanlon, who distinguishes “*structural* claims about rationality” from “*substantive* claims about what is a reason for what” (2003, p. 84). Scanlon (2003) doesn’t frame the distinction as a one between two kinds of rationality, but Scanlon (2007) comes close to framing it this way. It has since become common to do so, likely in part due to the increasing popularity of reasons-responsiveness accounts.

and Errol Lord (2018) have both argued that structural rationality is reducible to substantive rationality.^{4,5} Alex Worsnip (2018), on the other hand, argues that the two can come into conflict, which suggests that neither is reducible to the other.⁶ I'll attempt to stay neutral in such debates. All I want to commit to here is that there is such a thing as substantive rationality, and it is the analysans of reasons-responsiveness theories. Since I'll be setting structural rationality aside for the rest of the paper, I'll hereafter refer to substantive rationality simply as 'rationality.'

As several theorists have noted, rationality is a property that seems to be intimately tied to credit and criticism. Attributions of rationality and irrationality are by their nature ways of crediting and criticizing agents for their actions and attitudes. As Parfit puts it:

When we call some act 'rational', using this word in its ordinary, non-technical sense, we express the kind of praise or approval that we can also express with words like 'sensible', 'reasonable', 'intelligent', and 'smart'. We use the word 'irrational' to express the kind of criticism we express with words like 'senseless', 'stupid', 'idiotic' and 'crazy'. To express weaker criticisms of this kind, we can use the phrase 'less than fully rational' (2011, p. 33).

Let φ stand for any rationally evaluable action or attitude. When an agent φ s rationally, she thereby deserves a kind of credit for φ ing as she does. And when an agent φ s irrationally, she thereby deserves a kind of criticism for φ ing as she does. According to Lord, this connection to credit and criticism is the "most fundamental feature" of rationality (p. 4).

I agree that this connection to credit and criticism is at least *a* fundamental (if not *the most* fundamental) feature of rationality. So, it's a precondition for the success of reasons-responsiveness theories that they successfully explain this connection. And they can do so only if they rely on accounts of correctly

⁴ Strictly speaking, it is not clear whether Kieseewetter and Lord are reductionists or error theorists about structural rationality. Lord, in particular, writes as if it is the merely *apparent* phenomenon of structural rationality that's accounted for by his theory of substantive rationality. Kieseewetter's remarks seem to be more neutral between reductionism and eliminativism. What's more, it isn't always easy to understand what the exactly difference between reductionism and error theory amounts to in this context. I thank Alex Worsnip for pointing this out to me.

⁵ By contrast, Broome (2007, 2013) has argued that structural rationality is the only kind of rationality, and that the kinds of substantive requirements that reasons-responsiveness theorists talk about, while genuine requirements of reason, shouldn't be thought of as requirements of *rationality* at all. I will set Broome's arguments aside in this paper.

⁶ On this issue, see also Lasonen-Aarnio (forthcoming).

responding to reasons that capture the sense in which, when we correctly respond to reasons, we are thereby creditworthy for our success in acting or holding attitudes in accordance with our reasons.

Before moving on, it is important to attend to another distinction, between ex ante and ex post rationality. When φ ing is ex ante rational for an agent, it is rational for that agent to φ . But this doesn't entail that the agent φ s *rationally*. In fact, it doesn't entail that she φ s at all. When an agent's φ ing is ex post rational, it is not just rational for her to φ , but she actually φ s *rationally*. So, in short, the distinction between ex ante and ex post rationality is the distinction between its being rational for one to φ and one's actually φ ing rationally.

Reasons-responsiveness theories are supposed to account for both ex ante and ex post rationality – that is, they tell us not just what it is for attitudes and actions to be rational to hold or perform, but what it is to act or hold attitudes *rationally*. They hold that to correctly respond to one's reasons in φ ing is to φ rationally – to be ex post rational. For some φ to be ex ante rational is for it to be favored by some set of reasons such that, if the agent correctly responded to those reasons in φ ing, her φ ing would be ex post rational.

2.2 *Correctly responding to reasons*

Because my goal in this paper is to provide an account of correctly responding to reasons, I'll endeavor to provide as ecumenical as possible of a characterization at this point. Perhaps the easiest way to accomplish this is to discuss some cases that are obviously *not* cases of correctly responding to reasons.

But first, a note on the reasons themselves. Our subject matter is normative reasons – considerations that count in favor of or against φ ing. This is important because one goal of analyzing rationality is to account for its normativity. Since normative reasons are obviously normative, analyzing rationality in terms of normative reasons has the promise of straightforwardly accounting for the normativity of rationality.

Within the realm of normative reasons, however, there's considerable room for disagreement about which reasons are rationalizing reasons. Almost everyone agrees that only those reasons that are in the agent's epistemic ken are rationalizing reasons. But reasons-responsiveness theorists disagree

about whether these reasons are subjective reasons, or some subset of objective reasons.⁷ The distinction between subjective and objective reasons is itself fraught, and I will attempt to remain neutral about it. But almost everyone agrees on two related features of the distinction. First, objective reasons are facts, whereas subjective reasons can be merely apparent facts. Second, objective reasons don't depend for their existence on being epistemically accessible by the agent, whereas subjective reasons do. Pointing out these two relatively uncontroversial features of the distinction is all I will do to characterize it for now.⁸ This is because my goal is to develop an account of correctly responding to reasons without antecedently taking a stance on which reasons rationalize. However, the account I develop will turn out to have implications for that debate. I'll discuss those implications in §6.2.

Whichever the rationalizing reasons are, one can φ in accordance with those reasons without correctly responding to them. Not only can one respond *incorrectly* to them, but one can fail to respond to them at all.

Suppose the forecast is for rain tomorrow. This is a reason for me to believe that it will rain tomorrow, but I could come to believe that it will rain tomorrow in a way that has nothing to do with this reason. Imagine I consult a psychic, who tells me it will rain tomorrow, and it's on that basis that I come to believe it will rain tomorrow. I now have the belief favored by the reason <the forecast is for rain tomorrow>, but I don't respond to this reason in believing it.

I might also respond to that reason, but incorrectly. Imagine I come to believe that it will rain tomorrow, on the basis of the reason <the forecast is for rain tomorrow>. However, this is because I also (irrationally) believe that the Illuminati controls both the forecast and the weather, so whenever

⁷ Subjective reasons are sometimes referred to as *apparent* reasons instead. For a defense of the subjective reasons approach, see Sylvan (2015). For a defense of the objective reasons approach, see Lord (2018).

⁸ A further complication is that on some views, "possession" of a normative reason is a non-factive relation. See, for example, Comesaña and McGrath (2014).

the forecast is for rain it will rain. Despite the fact that I believe on the basis of a good reason, there seems to be something deficient about the way in which I do so.⁹

Finally, I might correctly respond to some of my reasons without its being the case that I correctly respond to my reasons overall. Imagine there are two highly reliable weather channels: one says it will rain tomorrow and the other says it won't. If I come to believe that it will rain on the basis of the first forecast and ignore the second, it seems natural to say I've correctly responded to one of my reasons, but I haven't correctly responded to my reasons overall, because I haven't responded at all to another reason I have. It's clear that if rationality consists in correctly responding to reasons, it must consist in correctly responding to reasons that are sufficient to rationalize the relevant response.

To correctly respond to one's reasons in φ ing, one must not only φ when one's reasons favor it; there needs to be some connection between one's φ ing and the fact that one's reasons favor it. Broome (2007) summarizes the point helpfully:

Even if you F whenever your reasons require you to F , you might not be responding correctly to reasons; it might just be a coincidence. Some appropriate connection must hold between your reasons and your F ing. It may need to be an explanatory one. Alternatively, a mere counterfactual connection may be enough. For instance, the necessary condition might be that you would not have F ed had your reasons required you not to F . (p. 351)

The possibility of incorrectly responding to a reason, rules out a simple but tempting account of correctly responding to reasons. On such an account, to correctly respond to some normative reason R to φ is just to φ on the basis of some motivating reason R and for R to also be a normative reason to φ . While this view is temptingly simple and straightforward, the Illuminati example above shows that it can't be right.^{10,11} Unfortunately, the task of providing an account of correctly responding to

⁹ The possibility of this way of responding to a reason, but incorrectly, depends on the common assumption that reasons are atomic facts. Some, such as Fogal (2016) argue that this common assumption is false. However, since the broader arguments of the paper do not depend on its truth, I will write as if it is true. But my official position on this issue is one of neutrality.

¹⁰ Again, this is on the assumption that reasons are atomic facts.

¹¹ For further argument that a simple conjunctive view can't be right, see Lord (2018, ch.6) and Lord and Sylvan (forthcoming).

reasons won't be so easy. It will require a search for the "appropriate connection" mentioned by Broome – a connection that rules out cases of mere coincidence between one's φ ing and the fact that one's reasons rationalized φ ing.

3. In Search of a Connection

In this section, I'll further discuss what the appropriate connection must look like, and consider a promising account, according to which the connection is a dispositional one.

3.1. Sensitivity to reasons

Consider again the case from above in which I respond to a reason, but incorrectly. I come to believe that it will rain tomorrow, on the basis of the reason <the forecast is for rain tomorrow>, because I (irrationally) believe that the Illuminati controls both the forecast and the weather, so whenever the forecast is for rain it will rain. Why is this a case of incorrectly responding to that reason?

A plausible answer is that this is a case of incorrectly responding to a reason because I lack sensitivity to what we might call the *contributory profile* of that reason.¹² The contributory profile of some reason R to φ specifies the normative support for φ ing provided by R across all contingencies. I lack sensitivity to the contributory profile of the reason <the forecast is for rain tomorrow> because my using it as a reason to believe that it will rain tomorrow is contingent on my irrational belief that the Illuminati controls both the forecast and the weather. Since the normative support that consideration provides for my belief is not actually contingent on its being the case that the Illuminati controls both the forecast and the weather, there is a mismatch between the reason's contributory profile and how I use it.

So, in order to correctly respond to reasons, we must use those reasons in ways that line up with their contributory profiles. Of course, it can't be a necessary condition that they line up completely, because many reasons will have contributory profiles that range over innumerable contingencies. To what degree they must line up is an interesting and difficult question, but I must set it aside here. For now, let's just say that we must use those reasons in ways that line up with contextually determined subsets

¹² I owe the term 'contributory profile' to Lord (in conversation).

of their contributory profiles. My hope is that I can rely on an intuitive grasp of such contextual relevance, and so proceed without an account thereof.

Why is it necessary for correctly responding to a reason that we be sensitive to its contributory profile? Because without such sensitivity, the connection between some consideration's being a reason to φ and our φ ing is too tenuous to count as an appropriate connection. This doesn't mean the appropriate connection is merely a counterfactual one. As Broome suggests, it is plausibly an explanatory connection. When there is an explanatory connection between some consideration's being a reason to φ and our φ ing, it seems like in an important sense no accident that we φ ed when that consideration was a reason to φ . So, it's plausible that that we correctly respond to some reason to φ only if we φ because it is a reason to φ . Moreover, it seems right to hold that when we fail to be sensitive to a reason's contributory profile, it is precisely this explanatory connection that is missing.

The idea that the appropriate connection is an explanatory one becomes even more compelling when we focus on the relationship between creditworthiness and reasons-responsiveness. If the notion of correctly responding to reasons is going to do the work in our theories that it promises to, it must be the case that when we correctly respond to reasons, we deserve credit for responding as we do. But it cannot simply be the truth of various counterfactuals that makes it the case that we're creditworthy when we correctly respond to reasons, because such counterfactual facts do not secure a relevantly non-accidental connection between our reasons and our responses.¹³ By contrast, when we φ *because of the fact* that our reasons favor φ ing, this does seem to secure the appropriate connection such that we deserve credit for φ ing in accordance with our reasons.

In light of this, I'll proceed on the assumption that the connection we're searching for is an explanatory connection between φ ing and R's being a reason to φ that makes φ ing sensitive to R's contributory profile. When such a connection obtains, it's no mere accident that we φ when our reasons favor φ ing. And, crucially, our φ ing in accordance with our reasons is non-accidental in a particular way, such that we deserve credit for it.

¹³ For an argument that the relevant sense of non-accidentality (or non-coincidence) cannot be merely counterfactual, see Faraci (2019). While Faraci's argument is about epistemic non-coincidence (the kind needed for knowledge), it applies more generally.

3.2. *The promise of dispositions?*

What sort of connection could obtain between our φ ing and R's being a reason to φ such that our φ ing is sensitive to R's contributory profile? There are many possible answers, but one that has gained considerable popularity recently is that it is a *dispositional* connection. Such a dispositional connection would consist in manifesting dispositions to φ when R is a reason to φ and to cease φ ing when R ceases to be a reason to φ . Defenders of this dispositionalist view maintain that when we manifest such dispositions, we φ in virtue of the fact that R is a reason to φ . Moreover, they maintain that manifesting such dispositions is sufficient to make our φ ing sensitive in the right way to R's contributory profile.

This is precisely the kind of analysis of correctly responding to reasons defended by Errol Lord in his account of (ex post) rationality. Because of this, I will use Lord's account as my primary foil in this paper. On his account, for some agent A to correctly respond to some reason R is "for A's φ ing to be a manifestation of A's knowledge about how to use R as the reason it is to φ " (p. 139). Lord understands the relevant know-how in terms of the sort of dispositions mentioned above, which he calls essentially normative dispositions. These are dispositions that are essentially sensitive to normative facts – in this case, facts about normative reasons.¹⁴

The dispositional account can explain why, in the case we began with, I incorrectly respond to the reason <the forecast is for rain tomorrow> in believing that it will rain tomorrow. My belief that it will rain tomorrow is a manifestation of certain dispositions I have to believe that proposition in certain situations. But these dispositions are insufficiently sensitive to the contributory profile of the reason for which I believe. So, although my belief is based on a consideration that constitutes a normative reason, it fails to be correctly responsive to that reason.

Manifesting reasons-sensitive dispositions secures a connection between our responses and the reasons for them that's not merely counterfactual. In fact, Lord argues, it secures a familiar type of explanatory connection: a causal one. When φ ing manifests knowledge about how to use R as the

¹⁴ Lord's language of "essentially normative dispositions" follows Wedgwood's (2006) language of "essentially rational dispositions). The dispositions in question are essentially normative in the sense that "it is part of the essence of these dispositions to be sensitive to normative features of facts" (Lord, 2019, p. 138). This notion may require further unpacking, but it's not fully clear how to do so.

reason it is to φ , the agent φ s *because* R is a reason to φ , and this is a causal sense of ‘because.’ Of course, the problem of deviant causal chains shows us that not just any causal connection is the appropriate kind for our purposes. However, because this causal connection occurs through the manifestation of a reasons-sensitive disposition, Lord contends that it’s exactly the right sort of explanatory connection.

Importantly, the dispositional account doesn’t require that the agent in any way takes the relevant consideration to be a reason for her to respond to it as a reason. As long as we manifest the right dispositions, we correctly respond to reasons, even if we have no mental states with normative content. In fact, on the dispositional account, we can correctly respond to reasons without in any way representing those reasons in our minds.

This means that, on many accounts of what it is to φ for a *motivating* reason, including Lord’s own, correctly responding to a normative reason doesn’t require φ ing on the basis of any motivating reason at all.¹⁵ This is because on such accounts, φ ing on the basis of a motivating reason involves φ ing in virtue of taking that reason to count in favor of φ ing. If we accept such accounts (as both Lord and I do), then the dispositional account entails a kind of disjunctivism about responding to reasons.¹⁶ According to this form of disjunctivism, responding to normative reasons and φ ing on the basis of motivating reasons are two distinct phenomena, and neither requires the other. This is a surprising conclusion, because despite the failure of the simple conjunctive account, we might have thought that φ ing on the basis of R, where R is a normative reason, was at least a *necessary* condition for correctly responding to R.

As it turns out, the lack of any representation condition in the dispositional account plays an important role for Lord, because it allows him to solve the New Evil Demon problem while holding onto the claim that rationalizing reasons are *objective* normative reasons. So, in addition to being an independently promising account of correctly responding to reasons, the dispositional account has particular explanatory power for objective reasons theorists. I will return to this issue in §6.2.

¹⁵ By motivating reasons, I simply mean those considerations on the basis of which we φ .

¹⁶ Because Lord accepts a representation condition on φ ing for motivating reasons, he embraces this kind of disjunctivism. For a defense of a representationalist account of motivating reasons that is neutral about responding to normative reasons, see Singh (2019).

The dispositional account is clearly promising. And Lord argues persuasively that when plugged into the reasons-responsiveness framework for analyzing rationality, it delivers correct verdicts in a variety of cases. But an important question remains in the discussion of the dispositional account: when we φ by manifesting reasons-sensitive dispositions, are we thereby creditworthy for φ ing? As discussed in §2.1, reasons-responsiveness theorists agree that rationality is intimately tied to creditworthiness. Since Lord emphasizes, and develops the dispositional account extensively, it's surprising that he largely takes it for granted that manifesting reasons-sensitive dispositions is sufficient to ground creditworthiness for our responses. It is precisely this assumption that I will question in the following section.

4. Disposing of Dispositionalism

In this section, I'll examine what is supposed to be special about reasons-sensitive dispositions such that they ground creditworthiness. I'll then argue that only dispositions with the right sort of basis in the agent can play this role. But if this is the case, the dispositions themselves cannot be what ground creditworthiness, so the dispositional account is inadequate.

4.1. *The appeal to competences*

Why think φ ing by manifesting reasons-sensitive dispositions makes us creditworthy for φ ing in accordance with those reasons? Lord says surprisingly little about this. One thing he does say is that manifesting reasons-sensitive dispositions is an achievement because such dispositions are *competences*. What does this mean? Here is Lord:

What type of competence is this [disposition]? It is a competence to correctly respond to reasons. That is, it is a competence that is sensitive to some of the facts about reasons. When one exercises this competence, one's actions are sensitive to the normative facts. One gets things *right*. Moreover, given the manifestation of a competence, it is no accident that one gets things right. This is why reacting for normative reasons is an achievement. And this is why one is creditworthy when one acts for the right reasons (p. 156).

This is somewhat mysterious. Lord claims that manifesting reasons-sensitive dispositions grounds creditworthiness because these dispositions are competences to correctly respond to reasons. But on the dispositional account, correctly responding to reasons consists in manifesting these dispositions. So, we are left with the claim that manifesting reasons-sensitive dispositions grounds creditworthiness

because these dispositions are competences to manifest these very dispositions. This leaves it entirely obscure what is special about *competences* such that manifesting them grounds creditworthiness.

To put the point another way, labelling reasons-sensitive dispositions competences just pushes back the question of creditworthiness. Now, instead of asking why manifesting reasons-sensitive dispositions grounds creditworthiness, we need to ask why manifesting competences grounds creditworthiness. Labelling certain dispositions ‘competences’ doesn’t help the dispositional account shed any light on the relationship between creditworthiness and correctly responding to reasons.

One thing that’s suggested by the passage above is that dispositions are competences when they dispose us to get things right. Indeed, Lord writes at various points as if this is what makes competences special. For example: “Importantly, this know how is a *competence*. It disposes one to get things right.” (p. 117). So, perhaps what’s special about competences is they reliably dispose us to success. This way of thinking about competence is in line with much of what Ernest Sosa, who popularized the language of competence in epistemology, says about what makes competences special. Sosa says that “a competence is a disposition, one with a basis resident in the competent agent, one that would in appropriately normal conditions ensure (or make highly likely) the success of any relevant performance issued by it” (2007, p. 29).¹⁷

If this is what competences are (and why they’re special), then the dispositional account has somewhat of a story about why reasons-sensitive dispositions ground creditworthiness. Reasons-sensitive dispositions to φ are competences because they reliably dispose the agent to achieve the relevant form of success, which in this case is φ ing in accordance with her reasons. This makes the appeal to competence somewhat less mysterious.

4.2. *Competences and creditworthiness*

If competences are dispositions to succeed, we’re left with the following question: does manifesting a disposition to succeed entail that the agent is creditworthy for her success? Proponents of the dispositional account think the answer is yes, but have done little to argue for this conclusion.

¹⁷ For further development of the notion of competence in his work, see Sosa (2010, 2015).

I think the answer is no – that is, I think manifesting any old disposition to succeed is *not* sufficient for the agent to be creditworthy for that success. This is because the manifestation of a disposition doesn't yield success that is attributable to the agent unless that disposition has a basis in the kind of categorical property of the agent that would make this disposition properly *hers* qua agent.¹⁸ Indeed, Sosa's own characterization of competence says that a competence is a disposition "with a basis resident in the competent agent." But for proponents of the dispositional account, this aspect of competence tends to drop out of the picture in favor of increased focus on reliability.

If competences are a special class of dispositions that ground creditworthiness, they cannot be just any old dispositions to succeed. They must be dispositions to succeed with the appropriate basis in the agent. Without such a basis in the agent, such dispositions would not be the agent's *qua* agent, and so success as a result of manifesting them could not be attributable to her. Manifesting such dispositions would not even be sufficient to make her actions, beliefs, and other responses intelligible, let alone rational.

This point is related to (though ultimately distinct from) one made by Warren Quinn (1993) with his oft-discussed 'radioman' case:

Suppose I am in a strange functional state that disposes me to turn on radios that I see to be turned off. Given the perception that a radio in my vicinity is off, I try, all other things being equal, to get it turned on. Does this state rationalize my choices? Told nothing more than this, one may certainly doubt that it does. But in the case I am imagining, this is all there is to the state (p. 236).

Quinn goes on to argue that the manifestation of this disposition "may help explain, causally, why I turn on a particular radio, but it does not make the act sensible" (p. 237). He is arguing here against a simple dispositional account of desire. On such an account, he argues, desires could not even make our behavior intelligible, let alone rational. And, as various discussions of Quinn's examples suggest, what is missing in this picture of desire is the agent's actually seeing something good or worth having

¹⁸ By 'attributable' I mean to pick out a property akin to responsibility in the attributability sense, as discussed by Shoemaker (2011), among others.

in the object of desire.¹⁹ Without such a component, there's nothing that makes the output of a desire attributable to the agent.

The dispositions Quinn discusses are, in an important sense, baseless dispositions. This isn't to say they're wholly metaphysically ungrounded; but since their basis is specified to be merely a functional state, they lack a basis in agent-level categorical properties. The dispositional account of correctly responding to reasons leaves room for the manifestation of similarly baseless dispositions, as long as those dispositions are sufficiently reliable. But it is a mistake to think that the addition of reliability makes any significant difference here. If it just so happened that one was reliably disposed to turn on radios only when doing so was independently desirable, this would not suffice to make the manifestation of that disposition any more intelligible of a behavior.

Consider the forecast case again. Suppose I have a disposition to believe that it will rain tomorrow when <the forecast is for rain tomorrow> is a reason for me to believe that. But this isn't because I in any way take what the forecast says to favor forming any particular beliefs. In fact, I have no views on the normative significance of the forecast. I simply have a disposition to form a certain belief when certain conditions obtain, with no basis in me *qua* agent. As such, I succeed in the sense that I believe in accordance with my reason; but I do not thereby deserve credit for my success.

Once we specify that the dispositions being manifested are baseless, it is no longer compelling that the agent is creditworthy in virtue of manifesting such dispositions. In fact, it seems to me that the dispositional account gains undeserved appeal from the fact that, when this is specification is not made, we are inclined to fill in examples by attributing a categorical property to the agent as a basis for the relevant disposition. This is why cases of manifesting reasons-sensitive dispositions seem to be cases where the agent is creditworthy for her success.

Examples involving baseless dispositions recall another famous example: Laurence Bonjour's (1980) case of Norman the clairvoyant, which provides a powerful objection to reliabilism about epistemic justification. This affinity isn't surprising, since the dispositional account ends up looking like a form of reliabilism about correctly responding to reasons. It wouldn't be difficult to construct clairvoyant-

¹⁹ For example, see Scanlon (1998, p. 38) and Railton (2012, p. 22).

style counterexamples to the dispositional account involving agents who manifest reliable dispositions to believe in accordance with normative reasons without actually basing their beliefs on those reasons.

The basic point here is this: when an agent manifests a baseless disposition to φ , her take on the reasons to φ play no role in her φ ing. As such, there is nothing that makes her φ ing intelligible from her own perspective. In general, when we respond in ways that are unintelligible to us, there's an important sense in which, if we respond successfully, this just *happens* to be the case. This is so even if in another sense it's no mere accident that we succeed when we manifest dispositions to succeed. The important sense of non-accidentality or non-coincidence here is whatever sense lines up with creditworthiness for success. And when we respond successfully even though our responses are unintelligible, it doesn't seem like we are creditworthy for our success.

This point can be further brought out by considering the influence of the agent's own take on reasons on our intuitions about creditworthiness. Take the following pair of cases:

May. May has promised a friend that she'll drive him to the airport, and that fact constitutes sufficient reason for May to drive him to the airport. May is disposed to drive her friend to the airport when that fact is a reason to do so, and her disposition has a basis in her recognition of that reason as such. Manifesting this disposition, she drives him to the airport.

Jay. Jay has also promised a friend that he'll drive him to the airport, and that fact similarly constitutes sufficient reason for Jay to drive him to the airport. Jay is disposed to drive his friend to the airport when that fact is a reason to do so. But this disposition has no basis in Jay *qua* agent. Jay doesn't recognize his reason as such, and instead takes the fact that his friend will owe him a favor to count in favor of his action. He also mistakenly thinks he's disposed to drive his friend to the airport when this other fact is a reason. Nevertheless, he manifests the disposition he actually has, and drives his friend to the airport.

May and Jay manifest the same disposition.²⁰ But it seems only May, and not Jay, deserves credit for acting in accordance with a reason. Unlike the clairvoyant, Jay's action is intelligible to him, because he has a story about why he did it. However, that story has nothing to do with the disposition that explains why he actually did it. This seems equally incompatible with his deserving credit for acting in

²⁰ I'm assuming that dispositions are not individuated by their bases. The point can easily be restated if this assumption turns out to be false.

accordance with a reason. And we can only explain why Jay lacks creditworthiness by appealing to something beyond the dispositions themselves.

Ram Neta (2019) uses similar cases to criticize dispositional accounts of the basing relation. As he points out, divergence between the agent's dispositions and her representations seems to prevent the manifestation of a disposition from being a justifying instance of the basing relation. This is something the dispositionalist can't explain. The same is true of the dispositional account of correctly responding to reasons. It cannot explain why dispositions to succeed fail to ground creditworthiness for success when they lack a basis in the agent.

5. The Return of Representation

Proponents of dispositional accounts of correctly responding to reasons tend to be reliabilists about competence. As such, they reject the idea that the agent's own representations need to play any role in correctly responding to reasons as overly intellectualized. My goal in this section is to first argue that this rejection is a mistake, and then, show how carving out a role for representations solves the problem of baseless dispositions.

5.1. Do representations overintellectualize?

In §4, I showed that reasons-sensitive dispositions can only ground creditworthiness if they have the right sort of basis in the agent. The most natural candidate for the appropriate basis is the agent's own normative outlook – roughly, how she represents the world, normatively. This includes, of course, her take on which responses are favored by which considerations. It's highly plausible that when manifesting reasons-sensitive dispositions makes us creditworthy for our responses, it is because those dispositions issue from our normative outlooks. In other words – it's plausible that having a basis in certain normative representations is what makes competences out of the relevant dispositions.

Something like this is suggested in Neta's account of the basing relation. While Neta doesn't explicitly discuss creditworthiness, he contends that justifying instances of the basing relation consist in manifesting a reasons-sensitive disposition that's partly constituted by the agent's representation of that very disposition as justifying. He then argues that this correctly rules out cases of dispositions that are misrepresented, or not represented at all, from giving rise to justifying instances of the basing relation. The question of what makes particular instances of the basing relation justifying is distinct

from, but closely related to, the question of what it is to correctly respond to a reason.²¹ While I won't be able to fully explore that connection here, Neta's view is worth noting because it shows that there are extant views on a related topic that recognize the role of representations in explaining the normative significance of certain dispositions.

Despite all of this, many theorists (especially proponents of dispositionalism) have been hostile to the idea that normative representations play a role in correctly responding to reasons. This is usually because they see it as an *overintellectualization*. Lord contends throughout his book that it would overintellectualize rationality to include a representation condition at any point in the analysis of correctly responding to reasons. He offers a few different arguments for this claim.

One argument is that a representation condition requires "one thought too many" (p. 103). For example, suppose I infer q from two known facts, p and $p \rightarrow q$. According to Lord, for this to be a case of correctly responding to p as a reason, all that needs to be the case is that I manifest my competence with modus ponens. It would be an overintellectualization to claim that I also need to believe that p is a reason to believe q .²²

I agree with Lord that the additional, explicit belief that p is a reason to believe q is unnecessary. But it's a mistake to assume that this is the only way representations can govern inferences. In fact, those who support a 'taking condition' on inference rarely spell that condition out in terms of additional, explicit beliefs about what particular token considerations are reasons for particular token responses.²³

Instead, it's much more plausible to think that more general representations are built into the relevant competence itself. It's true that to correctly respond to a reason by making a modus ponens inference, I need only manifest my competence with modus ponens. The issue is what counts as a competence with modus ponens. And as Paul Boghossian has argued in a number of papers, a disposition to make

²¹ The relationship between the two questions depends partly on the relationship between rationality and justification. See Sylvan (ms) for discussion.

²² Furthermore, it would generate a kind of Carrollian regress. I discuss the worry that representationalism generates regresses in §5.2.

²³ The term 'taking condition' is has bene popularized by Paul Boghossian's work on inference. See, for example, Boghossian (2014).

transitions in accordance with modus ponens is not enough. As I've argued, such dispositions must have a basis in the agent. This basis could be a general representation of the modus ponens schema, rather than a representation of normative support between a particular reason and a particular belief. So, the involvement of representations does not introduce, as Lord claims it does, a proliferation of representations specific to token instances of inferences.

This point applies to the competences involved in reasons-responsiveness more generally. To claim that competences are governed by representations is not to claim that each instance of responding to reasons is governed by its own discrete representation. Just as multiple instances of responding to reasons might manifest the same disposition (whose manifestation-conditions might correspond to a more general principle), these competence-constituting dispositions can also be governed by representations of more general schemata (which themselves might correspond to the same general principles). And these representations don't need to take the form of explicit beliefs to play this role.

The same sort of point can be made in response to Lord's second argument, which is that when we look at other competences, such as linguistic competence, it's not plausible that they require representations. For example, it's not plausible that manifesting competence with the word 'red' requires having "beliefs about the semantic relations between objects with a certain hue and the word 'red'" (p. 119).

Again, I agree with Lord that such beliefs are unnecessary. But again, it's a mistake to think this is the only way in which representations could be necessary for competence. In fact, many linguists and philosophers of language who study linguistic competence hold that there is a representational component to it. For example, Chomsky (1980) argues that linguistic competence is not characterizable solely in terms of dispositions, capacities, or abilities precisely because it constitutively involves representations. These representations consist not in explicit beliefs about particular objects and words, but rather mental structures that encode more general linguistic patterns.

Of course, representationalist theories of linguistic competence are controversial. But the fact that they are considered plausible by many linguists and philosophers of language makes linguistic competence an unsuitable example if the goal is to show that competences in general don't require representation. If competences are dispositions governed by representations, they are governed by representations of patterns, rules, or schemata, not by the kinds of beliefs discussed by Lord.

Finally, Lord argues that representation conditions are simply too demanding. Here, he quotes Parfit:

We can have rational beliefs and desires, and act rationally, without having any beliefs about reasons. Young children respond rationally to certain reasons or apparent reasons, though they do not yet have the concept of a reason. Dogs, cats, and some other animals respond to some kinds of reason...though they will never have the concept of a reason. And some rational adults seem to lack this concept (2011, p. 118).

Lord uses this quote to bring out the implausibility of a view he calls the Reasons Belief view, according to which correctly responding to a reason requires believing the relevant consideration to be a reason to φ (p. 102). According to Lord, this view is implausible because, since beliefs represent their contents, it entails that creatures that lack the concept of a reason can possess reasons.

There are two responses to this argument. The first is that talk of ‘beliefs about reasons’ evokes explicit beliefs about what particular token considerations are reasons for particular token responses. And it’s indeed implausible that such beliefs are necessary for correctly responding to reasons. But as I’ve already pointed out in this section, this isn’t what the representationalist should think anyway. Instead, the representationalist should think that more general representations come into the picture by constituting the categorical basis of reasons-sensitive dispositions. In other words, the agent’s representations are those of her categorical properties in virtue of which she is disposed to φ when she has sufficient reason to φ .

The second response is that both Parfit and Lord, as well as others who complain about overintellectualization, drastically overstate the case for including young children and non-human animals as creatures that respond to reasons in the same sense that we do. Notice that when Parfit and Lord complain about overintellectualization, talk of creditworthiness drops out. But for both of them, the connection to credit and criticism is supposed to be fundamental to rationality. Indeed, it was their own remarks I appealed to in connecting rationality to creditworthiness in §2.1. This generates an internal tension for Parfit and Lord.

We generally don’t treat young children and non-human animals as the sort of rational agents who are the appropriate objects of credit and criticism. We don’t hold them responsible in the way we hold full rational agents responsible, and we don’t consider their success or failure to be attributable to their agency. So, if rationality, and thereby responsiveness to reasons, are so deeply connected to credit and criticism, why should we think young children and non-human animals are capable of responding to

reasons in the way full rational agents are? Parfit and Lord's own commitments undermine their complaints.

Of course, there is *some sense* in which it's right to say that young children and non-human animals respond to reasons. But whatever this sense is, it's much more minimal than whatever is required for rationality, conceived of as an achievement of agents who are fully appropriate objects of credit and criticism. This points to a way in which the dispositional account might actually *underintellectualize* rationality. Dispositionalism allows for beings that are not appropriate objects of credit or criticism to count as correctly responding to reasons. This violates a shared commitment among reasons-responsiveness theorists: that rationality and creditworthiness are deeply and necessarily connected. So, reasons-responsiveness theorists must abandon their anti-intellectualism.

5.2. Reasons-responsiveness requires representation

Proponents of the dispositional account of correctly responding to reasons eschew representation because they think it overintellectualizes rationality. I've shown that this is a mistake. In doing so, I've also outlined the shape of an account that carves out an important role for representations. At a certain level of description, such an account looks like the dispositional account. That's because I'm happy to agree with Lord that to correctly respond to reasons is to manifest a competence. I'm also happy to agree that this competence involves reasons-sensitive dispositions.

But I disagree with the dispositional account's commitment that the reliability of such dispositions is sufficient to make them competences. No matter how reliably they dispose us toward success, dispositions are not normatively significant unless they are *our* dispositions. Without an agent-level basis, the manifestation of a disposition cannot be an expression of our agency, and thus cannot be to our credit. The missing piece, I have claimed, is the agent's own normative outlook.

What does this look like? I've gestured at some examples already. Being competent with modus ponens, I've suggested, requires not just that the agent be disposed to make transitions in accordance with modus ponens, but also that this disposition be governed by a representation of the modus ponens schema or inference pattern. Again, this should not be confused with a requirement that the agent think thoughts like "*p* is a reason to believe *q*" or "*q* follows from *p*" while drawing the conclusion.

We can generalize from this example. Recall that according to the dispositional account, an agent correctly responds to a reason when she manifests a disposition that's sufficiently sensitive to its contributory profile. When q follows from p , p is a reason to believe q that has a particular contributory profile. To be disposed to make transitions in accordance with modus ponens just is to have a disposition to believe q that's sensitive to this reason's contributory profile. What my account adds is that this disposition must be sensitive to the reason's contributory profile in virtue of having a basis in the agent's representation of that contributory profile. In the modus ponens example, a representation of the modus ponens schema plays this role.

In brief, this gets us the following view:

Representationalism: A correctly responds to some reason R in φ ing if and only if A manifests a disposition to φ that is sufficiently sensitive to R 's contributory profile in virtue of having as its categorical basis A 's representation of R 's contributory profile.

A representation of a reason's contributory profile is a representation of a consideration's normative support for a response under various contingencies. So, the representations involved in correctly responding to reasons are normative representations – that is, they are representations with normative content. So, I embrace the view that correctly responding to reasons requires possessing the concept of a reason.²⁴ In fact, it's precisely this commitment that allows my account to explain the relationship between creditworthiness and correctly responding to reasons – something the dispositional account has failed to do.

Consider **May** and **Jay** again. May and Jay manifest the same disposition, but it seems only May, not Jay, deserves credit for acting in accordance with a reason. My account makes good sense of this. While both May and Jay manifest the requisite reasons-sensitive disposition, Jay's disposition lacks the appropriate categorical basis because he fails to have a sufficiently accurate representation of the reason's contributory profile. By contrast, May's disposition has a basis in precisely such a

²⁴ In other work, I've argued that φ ing on the basis of motivating reasons requires possessing the concept of a normative reason, whether or not the motivating reasons on the basis of which one φ s are normative reasons (see Singh, 2019). If this is right, then correctly responding to normative reasons requires the same concepts as the more general case of φ ing on the basis of motivating reasons.

representation, and thus a basis in her *qua* agent. In other words, while May manifests a competence, Jay manifests a mere disposition. And mere dispositions lack normative significance.

Why does the addition of a normative representation make May creditworthy where Jay is not? The answer is that the kind of representation I have in mind is a *committal* representation: in representing some consideration as a reason, May commits herself to its being a reason. And it is precisely in virtue of this commitment that May has and manifests a disposition that's sensitive to that reason. Unlike Jay, May implicates her agency in her dispositions. This is why, when she manifests that disposition, her success in acting in accordance with a reason is attributable to her. So, this is why she deserves credit for her success while Jay does not.²⁵

At this point, proponents of the dispositional account might level another objection against representationalism. So far, I've only claimed that the representations involved in correctly responding to reasons must be sufficiently accurate. But it might be objected that this is insufficient, because in order for these representations to confer a positive normative status like creditworthiness on agents, they must themselves enjoy some positive normative status. In fact, one might think, in order for representations to confer creditworthiness on us for our responses, we must be creditworthy for the accuracy of those representations. This raises the threat of a regress of creditworthiness similar to the regresses of justification that have long been discussed in epistemology.

I'm sympathetic to the thought that the representations involved in correctly responding to reasons must themselves have some positive normative status, so I take the threat of regress seriously. Fortunately, however, it's plausible that the correct solution to the regress problem in epistemology will be generalizable to a solution to the threat of regress here.²⁶ To show this, it will be useful to briefly consider what foundationalist and coherentist solutions might look like when applied my account.²⁷

²⁵ This is, of course, only a brief characterization of what it is for a representation to be committal. Unfortunately, it is beyond the scope of this paper to provide an account of commitments. For a more detailed characterization of commitments that is congenial to what I say in this paper, see Neta (2018).

²⁶ This is no coincidence: doxastic justification and ex post rationality are very closely related properties.

²⁷ For a canonical discussion of foundationalism and coherentism in epistemology, see Sosa (1980).

A foundationalist solution would distinguish between basic and non-basic normative representations. If an instance of correctly responding to reasons involves a non-basic normative representation, this representation must be not just accurate, but itself correctly responsive to reasons. By contrast, if an instance of correctly responding to reasons involves a basic representation (such as a perceptual representation), this representation need not itself be correctly responsive to reasons. Instead, it must simply be a representation to which the agent is entitled, where entitlement is a distinct property from rationality that does not require responsiveness to reasons.²⁸ Since perceptual representations, among others, are paradigmatic candidates for representations to which we are entitled, this would provide a plausible foundationalist picture of how normative representations can play the role they need to in correct responsiveness to reasons.²⁹

A coherentist solution would instead require that the agent's normative representations fit together in the right way. Considering the myriad relations various reasons have been theorized to stand in (enabling and disabling, intensifying, etc.), it's easy to see how representations of various reasons' contributory profiles could either form a coherent overall normative outlook, or come into conflict by representing the world in incompatible ways. On the coherentist solution, these representations have the relevant positive status in virtue of fitting into a coherent web of representations that constitute the agent's overall normative outlook.

While I'm more partial to foundationalism than coherentism, I don't need to decide between them here. My point is not that there is an obviously correct solution to the regress problem. Rather, my point is that it's a new version of an old problem, and there is plenty of reason for optimism that whatever the solution to the old problem is, it can be generalized to provide a solution to the new one. So, while I grant that the representations involved in correctly responding to reasons must themselves have some positive normative status, I don't think that compels me to solve the regress problem here.

²⁸ The fact that basic representations aren't themselves responsive to reasons doesn't imply that they are somehow blind or unintelligent. A representation that isn't responsive to reasons can still be dynamic in the sense that it updates in the face of certain kind of feedback. The idea of states that are intelligent and can confer positive normative status upon further responses without themselves being responsive to reasons is discussed at great length in the recent work of Peter Railton, especially in his 2018 Locke Lectures. See also Railton (2012, 2017a, 2017b).

²⁹ The notion of entitlement figures prominently in recent foundationalist thought. See, for example, Burge (1993), Peacocke (2004) and Wright (2004).

Here's what I've done so far: I've shown that the dispositional account is inadequate because it cannot explain the element of creditworthiness in correctly responding to reasons. In doing so, I have motivated the need for a further condition on correctly responding to reasons. Finally, I've sketched that further condition in the form of representationalism about correctly responding to reasons.

Representationalism rejects the claim that manifesting reasons-sensitive dispositions is sufficient for correctly responding to reasons. But while representationalism is in this way incompatible with the dispositional account, it still leaves room for an important role for dispositions as expressions of the commitments that come out of our normative representations. In slogan form: to correctly respond to reasons is to manifest reasons-sensitive dispositions in virtue of representing those reasons as the reasons they are. Both dispositions and representations play an important role, but it is fundamentally the agent's representations that confer normative significance on the disposition-manifestations that constitute instances of correctly responding to reasons. So, it is ultimately a representationalist account.

6. Implications

This section will conclude the paper by discussing two implications of representationalism. First, representationalism undermines the motivation for disjunctivism about responding to reasons. Second, representationalism shows that objective reasons versions of reasons-responsiveness theories of rationality cannot solve the New Evil Demon problem.

6.1. *Disjunctivism*

According to the dispositional account, correctly responding to normative reasons is a matter of manifesting dispositions that are sensitive to those reasons. According to the view of motivating reasons that both Lord and I accept, φ ing on the basis of motivating reasons is a matter of responding in virtue of taking them to be normative reasons, whether or not they are. As mentioned in §3.2, this commits Lord to a kind of disjunctivism about responding to reasons, according to which φ ing for a motivating reason is not a necessary condition for correctly responding to a normative reason.

This is a surprising conclusion. As Lord points out, it's widely held that responding to a normative reason is a special case of φ ing for a motivating reason. But Lord thinks this is false, because while φ ing for a motivating reason requires representing it as a normative reason, responding to a normative

reason doesn't require any representation at all. So, for Lord, responding to normative reasons is not a special case of responding to motivating reasons, but instead a distinct phenomenon. Nevertheless, he claims, both of these phenomena are members of a higher-order disjunctive kind.

The kind of disjunctivism defended by Lord comes at a theoretical cost, because unity is a theoretical virtue, and disjunctive kinds are not truly unifying. But Lord thinks that since we have independent reason to accept a representation condition on φ ing for motivating reasons, but reject it when it comes to responding to normative reasons, we have independent reason to accept disjunctivism.

If we accept representationalism about responding to normative reasons, on the other hand, we aren't at all compelled to accept disjunctivism. On the contrary, it gives us a principled argument for the view that responding to normative reasons is a special case of φ ing for motivating reasons. Assume that Lord and I are right about motivating reasons; to φ on the basis of a motivating reason is to φ in virtue of representing it as a normative reason. According to representationalism, to correctly respond to a normative reason in φ ing is to φ in virtue of representing that reason's contributory profile (where this representation is sufficiently accurate and itself has the requisite positive normative status).

Representationalism entails that when one responds to some reason R as a normative reason, one also φ s on the basis of R as a motivating reason. This is because in both cases, one represents R as a normative reason. R's being one's motivating reason is strictly weaker in two ways: first, it doesn't require that one actually φ s in accordance with a normative reason, and second, it doesn't require one's representation of R as a normative reason to be accurate, or to have any positive normative status.

It's worth reiterating here that representationalism is *not* what Lord calls the "composite account," according to which responding to a normative reason is just φ ing on the basis of a motivating reason that happens to be a normative reason. Such simple conjunctive accounts were rejected at the outset of this paper. According to Lord, the failure of the composite account motivates disjunctivism. This is false, because disjunctivism and the composite account are not the only options. As I've just shown, representationalism entails that responding to a normative reason is a special case of φ ing on the basis of a motivating reason, but does *not* entail the composite account.

There's one more thing to say here. Lord also motivates disjunctivism using another putative contrast between φ ing for motivating reasons and responding to normative reasons, which is that the former

makes one's responses intelligible, whereas the latter makes them *achievements*. As Lord points out, mere intelligibility is not an achievement.

I agree with all of this, but I think it tells *against* disjunctivism rather than in favor of it. Lord is obviously right to point out that intelligibility is not sufficient for achievement. But it's quite plausible that intelligibility is *necessary* for achievement. As I argued in §4.2, it's difficult to see how one could deserve credit for φ ing if one's φ ing is not even intelligible. So, if φ ing for motivating reasons is what makes one's φ ing intelligible, as Lord and I both think, we should think φ ing for motivating reasons is a precondition for responding to normative reasons. In other words, we should think that responding to normative reasons is a special case of φ ing on the basis of motivating reasons. Once we notice that intelligibility is a precondition for creditworthiness, this further undermines disjunctivism instead of supporting it.

6.2. *Objective reasons theories*

As I mentioned in §2.2, reasons-responsiveness theorists disagree about whether rationalizing reasons are subjective reasons, or some subset of objective reasons. I have tried to develop an account of correctly responding to reasons without antecedently taking a stance on which reasons rationalize. None of the arguments I offered in favor of representationalism turned on whether rationalizing reasons are objective or subjective reasons.

However, representationalism turns out to have substantial implications for that debate. This is because the debate is a version of the debate between externalism and internalism about rationality. And externalist theories of rationality face the New Evil Demon problem. As it turns out, Lord's solution to the New Evil Demon problem depends on dispositionalism about correctly responding to reasons. I've argued on independent grounds that we should reject dispositionalism in favor of representationalism. As I'll show, this undermines Lord's solution to the New Evil Demon problem, and thereby undermines objective reasons theories.

Here is a brief explanation of the New Evil Demon problem, how it applies to objective reasons theories, and Lord's solution.³⁰ The problem is as follows: imagine that in two possible worlds, w_1 and w_2 , you have the same experiences and apparent memories, and you reason in the same way to arrive at the same beliefs, actions, and other reactions. It seems obvious that in each world, your reactions are equally rational.

Now, imagine that while in w_1 you're getting things mostly right, in w_2 you are being systematically deceived by an evil demon. As Wedgwood (2002) notes, intuitively, it's still the case that you're equally rational in both worlds. Internal duplicates – agents with identical non-factive mental states – seem to be equally rational. This suggests that rationality (at least) supervenes on agents' non-factive mental states. But externalists about rationality hold that rationality *depends* on something other than agents' non-factive mental states.

Here's how the New Evil Demon problem applies to objective reasons theories of rationality like Lord's. According to such theories, rationality depends on (some subset of) objective normative reasons, which are facts. But the relevant facts exist only in the good cases. The worry is that if the rationalizing reasons present in the good cases are missing in the bad ones, then objective reasons theories yield the incorrect verdict that bad case agents are irrational despite being internal duplicates.

For example, consider the following pair of cases. In the good case, it's cold outside and this fact is an objective reason for you to wear a jacket. Correctly responding to this reason, you wear a jacket when you go outside. It seems you are rational. In the bad case, it appears to be cold outside, but you're being deceived by an evil demon. Reasoning in exactly the same way, you wear a jacket when you go outside. It seems you're no less rational in the bad case than in the good case. However, in the bad case, you can't be correctly responding to the objective reason that it's cold outside, since that's not a fact. The burden is on objective reasons theorists to explain why you're still rational in the bad case.

³⁰ The New Evil Demon problem was originally raised by Lehrer and Cohen (1983) as a problem for reliabilism about epistemic justification. It has since become a sticking point in the related debate between internalists and externalists about the rationality of belief and other reactions (see especially Wedgwood (2002)).

This ‘backup reasons’ strategy says that for every objective reason of the form $\langle p \rangle$ in good cases, there are backup reasons roughly of the form $\langle \text{it appears that } p \rangle$ in bad cases. Furthermore, whenever the non-backup reasons are sufficient rationalizers, so are the backup reasons. This guarantees that if there are sufficient rationalizers in the good case, there are sufficient rationalizers in the bad case too (where sufficient rationalizers are objective reasons that are sufficient to make a response rational). Although I’m suspicious of this claim, I’ll grant it to Lord for the sake of argument.³¹

Lord’s work isn’t done, however. To solve the New Evil Demon problem for *ex post* rationality, it’s not enough to show that if there are sufficient rationalizers in the good case, there are sufficient rationalizers in the bad case. It must also be shown that in the bad case, the agent actually responds *rationally* – that she correctly responds to her reasons. So, if the backup reasons strategy is going to work, it must be that in the bad case, the agent correctly responds to the backup reasons.

This is where dispositionalism comes in. On the dispositional account, correctly responding to reasons is a matter of manifesting reasons-sensitive dispositions. And as Lord argues, it’s plausible that when one manifests a disposition to φ that’s sensitive to a reason of the form $\langle p \rangle$, one also manifests a disposition that’s sensitive to a reason of the form $\langle \text{it appears that } p \rangle$. For example, in the good case, you manifest a disposition to wear a jacket when the fact that it’s cold outside is an objective reason to do so. You also thereby manifest a disposition to wear a jacket when the fact that it appears to be cold outside is an objective reason to do so. In the bad case, you only manifest the latter disposition.

Thus, the dispositional account yields the verdict that in the bad case, the agent correctly responds to the backup reasons. But the dispositional account is false. Manifesting reasons-sensitive dispositions is insufficient for correctly responding to reasons. Correctly responding to reasons also requires that those dispositions have a categorical basis in the agent’s representations of reasons.

This spells trouble for objective reasons theories, because it looks like the agent’s responses are explained by the very same representations in both the good cases and the bad ones: representations of the *non-backup* reasons. Ordinarily, as Lord himself admits, it’s what appears to us as facts that we

³¹ For an argument against Lord’s claim about sufficient rationalizers, see Littlejohn (2019).

represent as reasons, not the fact that they appear to us.³² This is evidenced by the fact that when we cite our reasons, we ordinarily cite putative worldly facts, not appearance facts. So, once representationalism is on the table, it's not plausible that we respond to the backup reasons whenever we respond to the non-backup reasons.

Lord's response to the New Evil Demon problem is ingenious, and probably the best response an objective reasons theorist can give. But since representationalism, not the dispositionalism, is correct, Lord's response fails. This suggests that, insofar as the New Evil Demon problem is compelling, it's a fatal problem for objective reasons versions of reasons-responsiveness theories of rationality.

As such, the debate about what it is to correctly respond to a reason turns out to have substantial implications for the debate over which reasons are rationalizing reasons. Objective reasons theories face the New Evil Demon problem, and the best solution available to them depends on an account of correctly responding to reasons I've argued is inadequate. Since reasons-responsiveness theorists should accept representationalism about correctly responding to reasons, they should also think that rationality does not consist in correctly responding to objective normative reasons. Instead, rationality consists in correctly responding to some non-factive kind of normative reason, whether we call them subjective reasons or something else.

³² Objective reasons theorists might respond that we represent both reasons in the good case. But this would mean that what is shared between the good case and the bad case is the representation of the backup reason, rendering the non-backup reason epiphenomenal. This would be at odds with the spirit of the objective reasons view, if not the letter of it.

References

- Arpaly, Nomy (2002). Moral Worth. *Journal of Philosophy* 99 (5):223-245.
- Boghossian, Paul (2014). What is inference? *Philosophical Studies* 169 (1):1-18.
- BonJour, Laurence (1980). Externalist Theories of Empirical Knowledge. *Midwest Studies in Philosophy* 5 (1):53-73.
- Broome, John (2007). Does Rationality Consist in Responding Correctly to Reasons? *Journal of Moral Philosophy* 4 (3):349-374.
- (2013). *Rationality Through Reasoning*. Wiley-Blackwell.
- Burge, Tyler (1993). “Content Preservation.” *The Philosophical Review*, 102 (4): 457-488.
- Comesaña, Juan & McGrath, Matthew (2014). Having False Reasons. In Clayton Littlejohn & John Turri (eds.), *Epistemic Norms*. Oxford University Press. pp. 59-80.
- Faraci, David (2019). Groundwork for an Explanationist Account of Epistemic Coincidence. *Philosophers' Imprint* 19.
- Fogal, Daniel (2016). Reasons, Reason, and Context. In Errol Lord & Barry Maguire (eds.), *Weighing Reasons*. Oxford University Press.
- Kauppinen, Antti (forthcoming). Rationality as the Rule of Reason. *Noûs*.
- Kiesewetter, Benjamin (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.
- Lasonen-Aarnio, Maria (forthcoming). Enkrasia or Evidentialism? Learning to love mismatch. *Philosophical Studies*.
- Lehrer, Keith & Cohen, Stewart (1983). Justification, truth, and coherence. *Synthese* 55 (2): 191-207.
- Littlejohn, Clayton (2019). Being More Realistic About Reasons: On Rationality and Reasons Perspectivism. *Philosophy and Phenomenological Research*.
- Lord, Errol (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.
- Markovits, Julia (2010). Acting for the right reasons. *The Philosophical Review* 119 (2):201-242.
- Neta, Ram (2019). The Basing Relation. *Philosophical Review* 128 (2):179-217.
- (2018). Rationally determinable conditions. *Philosophical Issues* 28 (1):289-299.
- Peacocke, Christopher (2004). *The Realm of Reason*. Oxford: Oxford University Press.

- Parfit, Derek (2011). *On What Matters*. Oxford University Press.
- Quinn, Warren (1994). *Morality and Action*. Cambridge University Press.
- Railton, Peter (2012). That Obscure Object, Desire. *Proceedings and Addresses of the American Philosophical Association* 86 (2):22-46.
- (2017a). At the Core of Our Capacity to Act for a Reason: The Affective System and Evaluative Model-Based Learning and Control. *Emotion Review* 9 (4):335-342.
- (2017b). Author Reply: Affect, Value, Uncertainty, and Action. *Emotion Review* 9 (4):354-355.
- (2018). Learning and Doing: Toward a Unified Account of Rationality in Belief, Desire and Action. *The John Locke Lectures*. Oxford University.
- Scanlon, Thomas (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- (2003). Metaphysics and Morals. *Proceedings and Addresses of the American Philosophical Association* 77 (2):7-22.
- (2007). Structural Irrationality. In Geoffrey Brennan, Robert Goodin, Frank Jackson & Michael Smith (eds.), *Common Minds: Themes from the Philosophy of Philip Pettit*. Clarendon Press.
- Shoemaker, David (2011). Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. *Ethics* 121 (3):602-632.
- Singh, Keshav (2019). Acting and Believing Under the Guise of Normative Reasons. *Philosophy and Phenomenological Research*. 99 (2): pp. 409-430.
- (forthcoming). Moral Worth, Credit, and Non-accidentality. *Oxford Studies in Normative Ethics*, vol. 10.
- Sosa, Ernest (1980). The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge. *Midwest Studies in Philosophy* 5 (1):3-26.
- (2007). *A Virtue Epistemology*. Oxford University Press.
- (2010). *Knowing Full Well*. Princeton University Press.
- (2015). *Judgment & Agency*. Oxford University Press UK.
- Sylvan, Kurt (2015). What apparent reasons appear to be. *Philosophical Studies* 172 (3):587-606.
- (manuscript). On Divorcing the Rational and the Justified in Epistemology.
- Sylvan, Kurt & Lord, Errol (forthcoming). Prime Time (for the Basing Relation). In J. Adam Carter & Patrick Bondy (eds.), *Well-Founded Belief: New Essays on the Basing Relation*.

Wedgwood, Ralph (2002). Internalism Explained. *Philosophy and Phenomenological Research* 65 (2):349-369.

——— (2006). The normative force of reasoning. *Noûs* 40 (4):660–686.

Worsnip, Alex (2018). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research* 96 (1): 3-44.

Wright, Crispin (2004). “Warrant for Nothing (and Foundations for Free)?” *Aristotelian Society Supplementary Volume*, 78 (1):167–212.