Rationality as Government By Reason

Antti Kauppinen (a.kauppinen@gmail.com)

Draft, April 18, 2016

It has been nearly a dogma of contemporary metanormative theory that reasons are one thing and rationality is another. Rationality is a matter of having a proper order in one’s mind, as John Broome (2013) puts it, and it seems your mind can be in proper order even if you fail to recognize what reasons there as a matter of fact are. For example, supposing that the armchair you’re sitting in is about to be set on fire, there might be a conclusive reason for you to stop reading right now, but even so, you’re not necessarily irrational if you continue. Yet, as many have lately observed, if rationality and reasons come apart, and it is our reasons that determine what we ought to do, why care about being rational at all? To solve this puzzle, some have proposed that from the first-person perspective, the demands of rationality and reasons coincide, because what rationality requires of us is determined by our own take on reasons. But this won’t suffice to capture why rationality matters unless we place some constraints on the accuracy of our perception of reasons. In this vein, I’ll defend a novel Aristotelian view, according to which for us to have proper order in our mind, our thoughts must be governed by our Reason, that is, our fallible competence to recognize what reasons there objectively are for us.

I’ll make the case in two steps. The first is a partial defense of the Scanlonian idea of rationality as self-governance – roughly, shaping our attitudes by our take on reasons. I provisionally endorse the thesis that part of having a proper order in our mind is having, or coming to have, the attitudes we take ourselves to have conclusive reason for. Broome (2013) and others have raised a number of serious challenges for such views. They include explaining why
rationality requires coherence when reason thoughts seem absent, and the seeming rarity of thoughts about reasons for attitudes. My answer begins with a thesis in the philosophy of mind that I call Normative Constitutivism: intentions and some beliefs are in part constituted by thoughts about reasons for further attitudes. (Scanlon himself flirts with this kind of view.) For example, a crucial difference between intentions and desires is that only intentions are essentially self-governing, which I maintain involves commitment to treating certain things as reasons for action. And believing that R is a reason for action, when one takes it to be up to oneself whether one acts, entails taking R to be a reason for intending, or so I argue.

Given Normative Constitutivism, it is possible to explain the irrationality of incoherent attitudes in terms of perceived reasons after all. For example, given the ubiquitous perceptions of reason that are constitutive of intention and beliefs about reasons for action, someone who fails to intend a believed necessary means for an end lacks an attitude she takes herself to have conclusive reason for. This kind of explanation of structural irrationality generates asymmetric diachronic requirements that are suitable for guiding the formation or retention of attitudes. Such requirements must be formulated carefully, however, so that they allow for later changes of mind while still providing the right kind of guidance.

While there’s more to say for self-governance views than many have been willing to grant, they nevertheless have implausible consequences. Most importantly, they allow bootstrapping random rational requirements into being merely by forming beliefs about reasons, however crazy they may be, and struggle to explain the normative force of charges of irrationality. These issues motivate my second main step, requiring some kind of responsiveness to objective reasons as a condition of rationality. Like Broome (2013), I believe that it can’t be a demand of rationality that we respond correctly to the reasons there are, or what would be reasons were our
beliefs true (Schroeder 2009), or reasons that we possess (Lord 2014). All these views classify us as irrational too easily, for example when we have well-founded but false beliefs about reasons. What we need to do is to strike the right kind of balance between the subjective aspect of rationality as having the attitudes we take ourselves to have reason to have, and the objective aspect of rationality as responsiveness to genuine reasons. I develop Kurt Sylvan’s (2015) proposal, which links “apparent reasons” to competence with objective reasons.

So my final thesis, Rationality as Government By Reason (RGR), is roughly that rationality requires S to F if and only if S competently takes herself to have conclusive reason to F. Like perceived reason views, RGR ties rationality to self-governance understood in terms of one’s take on reasons for attitudes. But requirements arise only when one’s perceptions of reasons arise from exercising a relevant competence, where something is a competence when it is a disposition to take something to be a reason if and only if it is a reason. Such dispositions are fallible: at least in unfavorable circumstances, we may competently take ourselves to have reason even if we don’t. So rationality isn’t assimilated to responsiveness to objective reasons. If we use the label “Reason” in a quasi-Aristotelian way to refer to the totality of our various fallible competences to recognize objective reasons, we can simply say that being rational is being governed by Reason. Given that the perceptions of reasons embedded in attitudes like intention competently track patterns of reasons, RGR also yields an explanation of structural irrationality when combined with Normative Constitutivism. Finally, if we don’t conform to the dictates of our Reason, even in the occasional case in which it leads us astray, we will be relying on unreliable dispositions or nothing at all, and will only get it right by chance. This accounts for the distinctive normativity of rationality: since we should at least try our best, we should exercise and listen to our Reason.
1. Rationality and Perceived Reasons

Let’s begin with the idea that rationality consists in having a proper order in one’s mind. What is involved in such a proper order? One obvious feature is that our attitudes must fit together in certain ways. For example, it seems we’re (locally) irrational if we intend to A and intend to B while knowing we can’t do both, or believe we ought to F and fail to intend to F, or if we intend the end without intending what we believe to be a necessary means that is up to us. Explaining such platitudes about structural rationality is a task for any account of rationality. Another important desideratum is explaining how it is rational for us to respond to different situations, such as having conflicting intentions. Perhaps most fundamentally, rationality governs diachronic transitions from one set of attitudes to another, not just synchronic combinations of attitudes. Rationality tells us how to arrive at a proper order in our mind, not just what it is to be in order.

If we begin with the synchronic aspect of rationality, we’re easily led to postulate a set of requirements of rationality that either tell us to have one attitude if we have another attitude, or to avoid certain combinations of attitudes. The first kind, or narrow-scope, requirements say, for example, that we’re rationally required to intend to F if we believe we ought to F. The second-kind, or wide-scope, requirements say that rationality requires us to either not believe that we ought to F or intend to F. Each kind of view faces significant challenges: narrow-scope views license bootstrapping rational requirements into being just by adopting any random attitude, possibly creating conflicting or mutually undermining demands (Broome 2013, Brunero 2010), while wide-scope views seem to permit wishful thinking (Schroeder 2009) and struggle to provide guidance to form any particular attitude (Kiesewetter 2013). Further, both kinds of view have problems with explaining the normativity of rationality, and require supplementation with
independent (and seemingly _ad hoc_) basing principles to provide guidance for rational transitions. I will not attempt to resolve these complex debates in this paper, but will rather focus on constructing a plausible alternative approach (although doing so turns out to require addressing some challenges for narrow-scope views).

Suppose, then, that we begin with the diachronic aspect instead. Perhaps rationality is fundamentally a matter of _imposing_ a proper order on our minds, where the standard for what is proper is given by our own take on reasons. The irrationality of having incoherent attitudes must then be explained away or shown to follow from governing ourselves by our own standards. As Joseph Raz (2005) puts it, the idea that coherence itself is a requirement of rationality is a “myth”. The alternative approach is most famously defended by Niko Kolodny (2005) and Thomas Scanlon (2007). One rationale for it is that what is distinctive of us as rational creatures is that we can reflect on our attitudes, ask whether we have sufficient reason for them, and modify them when they don’t match what we take reasons to require. It is a familiar quasi-Kantian thought that self-government consists in shaping judgment-sensitive attitudes like beliefs and intentions by our take on the reasons we have. Could this be all there is to rationality? Hardly, but I think there is more to be said for this approach than is generally assumed.

Let us begin with explicit synchronic and diachronic formulations of the account (cf. especially Kolodny 2005 and 2007):

_Rationality as Responsiveness to Perceived Reasons, Synchronic (RRPR)_

Rationality requires $S$ to have attitude $A$ if and only if $S$ takes herself to have conclusive reason to have $A$.

_Rationality as Responsiveness to Perceived Reasons, Diachronic (RRPR-D)_
If \( S \) at \( t_1 \) takes herself to have conclusive reason to have attitude \( A \) at or from \( t_2 \) (where \( t_2 \) is simultaneous or later than \( t_1 \)), \( S \) is rationally required to form and maintain \( A \) at or from \( t_2 \), unless she changes her verdict on reasons for \( A \) by \( t_2 \).

(For brevity, I’ll omit the corresponding negative formulations, which raise some complications.) Although the diachronic version is arguably more fundamental, it will be simpler to deal with some issues by beginning with a synchronic version of the account. It is a special kind of narrow-scope requirement of rationality. It says that the requirement to have an attitude results from a second-order attitude, a normative stance towards one’s own attitudes. It is thus plausibly a kind of \textit{self-governance requirement}: we govern ourselves by taking and responding to a reflective stance towards how we ought to be.

It is noteworthy that RRPR talks about \textit{taking oneself to have reason} rather than belief in reasons. This is in order to avoid the problem of overintellectualization: it seems that we don’t have enough beliefs about reasons to account for all instances of rationality, and creatures that lack the concept of a reason can plausibly be irrational. What is it to take oneself to have reason? On the view of normative reasons I have defended elsewhere (Author 2015), we respond correctly to a reason \( R \) for F-ing by thinking in an F-friendly way when doing something F-relevant, which includes the various ways we can be causally influenced in the direction of F-ing by thoughts of \( R \) – being motivated to F, assigning F-ing a higher utility in deliberation in the presence of \( R \), raising the deliberative salience of F, and acquiring a disposition to do any of the previous. It’s a natural extension of this view to say that \textit{taking oneself to have reason} is \textit{endorsing} such influence, which might consist in something as minimal as willingness to appeal to \( R \) to defend thinking in a F-friendly way, if challenged by others, or a disposition to keep
thinking in a F-friendly way if it becomes salient that I do so because of R, or a disposition to be puzzled or critical of others for failing to think in a F-friendly way when they’re aware of R. This doesn’t require having the concept of a reason – unless, of course, we take such dispositions themselves to be constitutive of concept-possession.

Could rationality consist simply in responding correctly to perceived reasons? John Broome (2007, 2013) vigorously denies this. I’ll discuss three specific challenges, starting with the Problem of Deriving Structural Rationality. By itself RRPR says nothing about structural rationality. Indeed, as Broome notes, even though RRPR is a self-governance requirement, it doesn’t by itself entail even rational requirements that seem to be closely related. These include the fact that necessarily, if you are rational then, if you believe your evidence shows that p, you believe p (which Broome labels the Evidential Condition), and that necessarily, if you are rational then, if you believe your reasons require you to G, you intend to G (the Enkratic Condition) (Broome 2013, 98). And that’s only the beginning, since any plausible account of rationality must also explain the irrationality of incoherent attitudes, such as failing to intend what one takes to be necessary means to one’s end.

Broome’s second challenge is what I’ll call the Problem of Independent Standards. It seems that rationality can require us to do something even if we don’t take ourselves to have reason for it – maybe we’re irrational in part precisely because we don’t take ourselves to have reason to do certain things. Here’s one of Broome’s examples: “There is obviously something irrational about a person who does not intend a means she believes is necessary to an end she intends, even if she herself does not believe her reasons require her to intend a means she believes is necessary to an end she intends.” (Broome 2013, 93) Finally, specifically against Scanlon’s account, Broome points out that people rarely have the kind of beliefs about reasons
Scanlon thinks are crucial for rationality. He says that, “The problem is that the antecedent in Scanlon’s condition, that you believe your object-given reasons require you to $F$, is rarely satisfied. People rarely have beliefs of this sort.” I’ll call this the Problem of Rare Application.

2. Normative Constitutivism

The common assumption behind Broome’s challenges to perceived reason views is that all reason-takings are optional. But there’s good reason to deny this assumption. The alternative is that taking oneself to have reason for further attitudes is partially constitutive of some of our attitudes. Call this alternative view Normative Constitutivism. If it’s true, it may be that all agents merely as believers and intenders perceive themselves to have reasons that rule out incoherent combinations.

I will restrict myself to three attitudes that are crucial for explaining structural rationality: belief in evidence, belief in reasons for action, and intention. Let us begin with belief in evidence, which features in what Broome labeled the Evidential Condition. Our question is: what is it to believe that $E$ is evidence for $p$? It is not to believe that $E$ obtains – you might well believe that $E$ without regarding it as evidence. Nor is it merely to believe that $E$ makes $p$ more likely. That’s just a belief about probability, not about evidence. You might think that the fact that you just spilled chocolate milk on your car seat makes it more likely that the car will smell bad tomorrow without yet thinking that it is evidence for the car smelling bad tomorrow. This is unsurprising, given that evidence is a normative rather than non-normative notion – as Jaegwon Kim puts it, “the concept of evidence is inseparable from that of justification” (Kim 1988, 390). Thomas Kelly (2014) goes so far as to suggest that insofar as evidence is what confers justification, the
terms ‘reason to believe’ and ‘evidence’ are roughly synonymous. What is belief in evidence, then? Here is my proposal for a partial analysis:

Belief in Evidence: for you to believe that $E$ is (sufficient) evidence that $p$ is, at least, for you to take $E$ to be a (conclusive) reason to believe that $p$, and to not believe that $\neg p$.

(Those who, unlike me, believe in wrong kind of reasons for attitudes, should add ‘object-given reason’ or ‘truth-related reason’ here.) That is, what is distinctive about believing something is evidence for something else is precisely taking it to provide reason for believing the other thing. When you believe that the fact that there’s smoke coming out of your neighbor’s garden is evidence that there’s fire in your neighbor’s garden, you take it that it is a reason to believe that there’s a fire in your neighbor’s garden.

Since my focus in this paper is on practical rationality, I will not make much of Belief in Evidence. But it suffices to illustrate how Normative Constitutivism will help RRPR account for platitudes about structural rationality. The Evidential Condition was that you’re necessarily irrational, if you believe your evidence shows that $p$, but fail to believe that $p$. According to Belief in Evidence, if you believe that $E$ is sufficient evidence for $p$, you take yourself to have conclusive reason to believe that $p$. Given RRPR, it follows that you are rationally required to believe that $p$. Hence, Belief in Evidence and RRPR together entail that anyone who believes her evidence shows that $p$ but fails to believe that $p$ is irrational. In other words, a normative analysis of what it is to believe that something is evidence for something else closes the explanatory gap that Broome alleges perceived reasons accounts suffer from. This will be a model for my response to the Problem of Deriving Structural Rationality. But notice that it also sets the stage for solving the Problems of Rare Application and Independent Standards. First, if people have
certain opinions about what attitudes they have reason to adopt merely in virtue of having beliefs about evidence and other attitudes, such opinions will not be rare. In particular, they will be there when needed to explain platitudes about structural rationality. Second, while people may optionally take themselves to have reason for various things, if the constitutivist line I’ve sketched is right, they are guaranteed to impose certain standards on themselves. These standards are independent from their particular take on reasons. If their optional perceptions of reasons conflict with those deriving from the very nature of their attitudes, they will be less than fully rational, which is what needed to be explained.

The next attitude I will focus on is belief in reasons for action. I will not attempt a full analysis of such beliefs here. What is of interest for our topic is whether they, too, are normatively constituted in the sense that they involve perceptions of downstream reasons. And indeed, I believe they do. To begin with, consider the oddity or even paradoxicality of saying “I have reason to F, but I don’t have reason to intend to F”. This makes sense only in the special context in which it is believed that you will F whether or not you intend to F, or that you won’t F whether you intend to or not – for short, when acting isn’t up to you (Broome 2013). But if your F-ing depends on your intending to F (and is recognized to depend on your intention by parties to the conversation), utterances of this sort will be infelicitous, at least. What explains this? My claim is that it is the tacit recognition of the fact that when you believe you have a reason for action that you take to be up to you, you already believe you have a reason for intending to act as well. A little more precisely, the thesis is the following:

**Action-Intention Principle**: if you believe that \( R \) is conclusive reason for you to \( F \), when you believe your \( F \)-ing is up to you, you take \( R \) to be a conclusive reason for you to intend to \( F \).
Note that the thesis is not that reasons for acting and reasons for intending necessarily coincide. In the Toxin Puzzle (Kavka 1981), for example, you may have a reason for intending that is not a reason for action. (Again, I’m skeptical of the existence of such ‘wrong kind of reasons’ for attitudes, as opposed to right kind of reasons for wanting to have an attitude, but I’ll leave this issue aside here.) My claim is just that when you take yourself to have reason to act and that your acting is up to you, you take yourself to have reason to intend, too. Here it may be tempting to add “insofar as you’re rational”. But this is not an option, if the project is to make use of the Action-Intention Principle in understanding rationality. The claim is a constitutive one: whether you’re rational or not, you just don’t think that you have conclusive reason to go to Paris while believing that it’s up to you whether you go, if you don’t think you have conclusive reason to intend to go to Paris. Here we might employ a test inspired by Moore’s Open Question Argument: someone who says “I have conclusive reason to go to Paris and it’s up to me whether I go or not, but do I have conclusive reason to intend to go to Paris?” manifests at best a partial grasp of the concept of a reason for action. Perhaps she is confusing a reason to go to Paris (a reason to act) with a reason to be in Paris, which isn’t similarly linked to a reason for intending.

The Action-Intention Principle is a crucial step in explaining why something like Broome’s Enkratic Condition holds. Roughly, when it is combined with RRPR, it follows that necessarily, an agent who believes that she has conclusive reason to F and that F-ing is up to her but fails to intend to F is not fully rational, since by virtue of her belief about conclusive reason for action she takes herself to have conclusive reason for an attitude she doesn’t have. Thus, like Belief in Evidence, the Action-Intention Principle together with RRPR offers a straightforward explanation of a basic (second-order) platitude about structural rationality.
To be sure, this is only the beginning of the story. How do we get from perceived reasons to the rationality of coherent attitudes? I will only focus on practical rationality here. In keeping with the constitutivist strategy, my explanation will draw on the nature of intention. To get started, let us ask what distinguishes intention from other pro-attitudes, such as desire. One difference is that intentions concern our own future (or current) actions, while desires range more broadly. But there are also first-personal, action-directed desires. Just as I can intend to jump in the water, I can desire to jump in the water. The key difference between such desires and intentions is that only the latter involve commitment to act. When we intend to act, we’ve in some sense made up our mind. My claim is that this commitment involves at least one kind of normative thought: that we have reason to do things that facilitate the realization of our aim, and to avoid doing things that stand in the way of its realization, including realizing incompatible goals. Here’s one way of spelling this out:

Intention Commitment: In intending to E, we take the fact that M-ing facilitates E-ing to be a reason of proportionate strength for us to M, and the perceived incompatibility of E-ing and F-ing as a conclusive reason not to F.

Note that Intention Commitment shouldn’t be mistaken for the implausible claim that whenever we intend to E, we take ourselves to have reason to E. We may certainly have akratic intentions, for example. It’s only downstream reasons that we take ourselves to have as part of intending something. Note also that this formulation leaves unspecified the relative times of the end- and means-actions; I’ll return to the issue in the next section.

My version of Normative Constitutivism is directly inspired by Thomas Scanlon’s (2007) view of structural rationality. There is, however, a very significant difference, which in my view
renders Scanlon’s insight unusable for a defense of RRPR. He initially considers a normative constitutivist thesis that having an intention to A at t involves being disposed to “take one’s doing A at t into account in further reasoning by, for example, treating the fact that doing B would facilitate doing A at t as a reason for doing B” (Scanlon 2007, 88). But he retreats from this claim to a much weaker thesis, according to which it is “irrational for someone who has decided to do A at t (and has not changed his or her mind about this) to refuse to treat the fact that B would facilitate this as a reason for doing B or to refuse to treat the fact that doing B would be incompatible with doing A at t as a reason against doing B.” (Scanlon 2007, 93, my emphasis) On Scanlon’s considered view, then, finding there to be a reason to take means to one’s end is not constitutive of intending, but something that is rationally required. But first, this amounts to giving up on his stated strategy of explaining structural irrationality in terms of what one considers oneself to have reason for. After all, what makes it irrational not to treat the (believed) fact that M-ing would facilitate E-ing as a reason to M, when one intends to E? The answer is not going to be that one takes oneself to have reason to treat the fact that M-ing would facilitate E-ing as a reason to M. (A regress looms this way.) Instead, Scanlon will have to appeal to an independent principle of rationality that requires one to treat the fact that some action would facilitate realizing one’s aims as a reason to perform it. RRPR only enters the stage when one is rational to the extent of perceiving the reasons one should perceive. This means giving up on the simple and appealing view that rationality amounts to having the attitudes one takes oneself reason to have.

Second, Scanlon’s view locates the irrationality in the wrong place. What is irrational is failing to (intend) to take the means while intending the end, not failing to think that facilitating the end is a reason to take the means. Scanlon objects to my kind of view by claiming that it is
possible but irrational to fail to “take one’s aim into account in the proper way in one’s subsequent reasoning”. He thus appears to think that something like Intention Commitment would rule out this sort of irrationality by definition. But since taking oneself to have a reason doesn’t guarantee one will reason correctly, Intention Commitment allows that we may fail to take our aims into account in the proper way in our subsequent reasoning. After all, you might think you have conclusive reason to take the means, but nevertheless fail to form the intention, or simultaneously think you have conclusive reason to do something incompatible.

3. Explaining Practical Rationality

In this section I’m going to argue that if the Normative Constitutivist theses I’ve just defended are true, a perceived reasons account like RRPR can after all meet the challenges that Broome raises. I will focus solely on structural features of practical rationality in this paper, as theoretical rationality involves additional issues to do with the nature of belief as such. I’ll also set aside vexed issues to do with the rationality of preferences and choices, although I think the present framework can be extended to deal with them.

3.1 Synchronic Rationality

I’ll first spell out how to account for the following platitudes about synchronic rationality without appealing to self-standing coherence requirements:

*Intention Inconsistency:* If you believe that you can’t both A and B, believe that whether you A or B is up to you, intend to A, and intend to B, you’re not fully rational.

*Instrumental Failure:* If you intend to E, believe that you will not E unless you M, believe that your M-ing is up to you, and not intend to M, you’re not fully rational.
Let’s begin with Intention Inconsistency. The explanation goes as follows:

1. S believes that she cannot both A and B and that whether she A-s or B-s is up to her, and intends to A and intends to B. (Intention Inconsistency)

2. If S intends to A and takes A-ing and B-ing to be incompatible, S takes the incompatibility of A-ing and B-ing to be a conclusive reason not to B. (By Intention Commitment.)

3. If S takes herself to have conclusive reason not to B while believing B-ing is up to her, S takes herself to have conclusive reason not to intend to B. (By the Action-Intention Principle)

4. If S intends to B and takes A-ing and B-ing to be incompatible, S takes the incompatibility of A-ing and B-ing to be a conclusive reason not to A. (By Intention Commitment.)

5. If S takes herself to have conclusive reason not to A while believing A-ing is up to her, S takes herself to have conclusive reason not to intend to A. (By the Action-Intention Principle)

6. So, S takes herself to have conclusive reason not to intend to A and conclusive reason not to intend to B. (From 1, 2, 3, 4, and 5)

7. Rationality requires S not to have an attitude if S takes herself to have conclusive reason not to have the attitude. (From RRPR)

8. So, rationality requires S not to intend to A and not to intend to B. (From 6 and 7)

9. So, S is irrational insofar as she both intends to A and intends to B, given her beliefs. (From 1 and 8)
Most premises of this explanation require little comment. It begins with describing an agent who is irrational, according to the Intention Consistency platitude. Premises 2 to 6 simply spell out what Intention Commitment and the Action-Intention Principle entail about such an agent’s take on reasons. Premise 7 is a plausible corollary of RRPR. As I formulated it, RRPR only talks about attitudes an agent is required to have, not about attitudes an agent is required to lack. I prefer to skirt the tricky issue of when attitudes are permitted as far as I can (though see the final section for some discussion). In particular, it’s not clear whether we’re rationally required not to have an attitude only if we take ourselves to have conclusive reason not to have it. (This would permit holding attitudes whenever we have no opinion about reasons for them, for example.) But it seems true that if we take ourselves to have conclusive reason not to have an attitude, we are rationally required not to have it.

Premise 8 draws the intermediate conclusion that both inconsistent intentions are in violation of a rational requirement. In other words, rationality requires that both intentions should go. I think this is a plausible implication of the account, since the incompatibility of intentions as such doesn’t tell us which if either intention we should rationally adopt. Of course, if S has further related thoughts, such as that she ought to intend to A but not B, rationality will require adopting the intention to A (by the Action-Intention Principle, assuming that ought-thoughts entail conclusive reason thoughts). Step 9 concludes the argument by observing that S is irrational given her intentions and beliefs. Since the argument doesn’t hang on the content of S’s intentions and beliefs, it shows that intending what one takes to incompatible ends is necessarily
irrational.\(^1\) Intention Inconsistency is thus explained without appealing to either narrow-scope or wide-scope coherence requirements.

How about instrumental rationality? Explaining the irrationality of instrumental failure in terms for reasons transmitted to means from the end (Schroeder 2009, Way 2010) has proven difficult, since it is rationally permissible for us to intend some end without having conclusive reason to do so, or thinking that we do. Indeed, it seems we’re less than fully rational if we don’t intend the means, even if we shouldn’t intend the end at all, and even if we think we shouldn’t intend the end (Wallace 2001). I believe this, too, can be explained in terms of Intention Commitment and the Action-Intention Principle. Focusing only on the simplest case of taking the means to be necessary for one’s end, the derivation goes as follows:

1. S intends to E, believes that M-ing is necessary in order to E and that M-ing is up to her, and does not intend to M. (Instrumental Failure)
2. If S intends to E and believes that M-ing is necessary in order to E, S takes the necessity of M-ing for E-ing to be a conclusive reason for her to M. (By Intention Commitment)
3. So, S takes herself to have conclusive reason to M. (From 1 and 2)
4. If S takes herself to have conclusive reason to M while believing that M-ing is up to her, S takes herself to have conclusive reason to intend to M. (By the Action-Intention Principle).
5. So, S takes herself to have conclusive reason to intend to M. (From 3 and 4)
6. Rationality requires S to have an attitude if and only if she takes herself to have conclusive reason to have it. (RRPR)

---

\(^1\) For this thesis, see Bratman 1987. It may need to be qualified, however – see O’Brien 2014 for persuasive counterexamples to the unrestricted version.

\(^2\) For complications, see Kolodny (forthcoming).

\(^3\) This is itself by no means unproblematic, since it means the diachronic wide-scope principle fails to provide any
7. So, rationality requires S to intend to M. (From 1, 5, and 6)
8. So, S is irrational. (From 1 and 7)

Again, the steps are very straightforward. Intention Commitment does a lot of work here, and when the means are believed to be necessary, simple steps lead to a stringent rational requirement to intend the means. I don’t claim that explaining instrumental rationality in general is simple – a lot of work clearly needs to be done to account for the rationality of taking the best non-necessary means.\(^2\) I will leave such issues for another occasion.

### 3.2 Diachronic Rationality

As Kolodny (2005, 517) points out, it is odd, at least, to talk about requirements for having a particular combination of attitudes rather than requirements to form or avoid attitudes. Rationality seems to primarily call for a certain response from us, rather than being in a certain way. If we respond in the right way, we’ll end up having the right attitudes. In this sense, diachronic requirements are more fundamental. Here, again, is the diachronic formulation of the perceived reasons view:

**Rationality as Responsiveness to Perceived Reasons, Diachronic (RRPR-D)**

If S at \(t_1\) takes herself to have conclusive reason to have attitude \(A\) at or from \(t_2\) (where \(t_2\) is simultaneous or later than \(t_1\)), S is rationally required to form and maintain \(A\) at or from \(t_2\), unless she changes her verdict on reasons for \(A\) by \(t_2\).

Requirements derived from this principle have a narrow scope: rationality doesn’t require a coherent combination of attitudes, but rather a specific attitude, in both cases an intention. This is

\(^2\) For complications, see Kolodny (forthcoming).
a good thing, since such a principle can, on the face of it, guide forming or losing attitudes, unlike a symmetrical requirement to avoid a combination of them. But the asymmetry also invites a challenge, since it seems that it can sometimes be rational to give up the attitude that involves a perception of reason rather than the consequent attitude. For example, suppose you intend to get to work in time, and realize that to do that, you must drive over a child playing on the street. According to a plausible diachronic specification of Intention Commitment, you take the necessity of driving over the child for getting to work in time as a conclusive reason to drive over the child now, so by the Action-Intention Principle and RRPR-D, it looks like you are rationally required to form an intention to do so. The wide-scope objection is, roughly, that it could also be rational to give up the intention to get to work in time.

Here the synchronic–diachronic distinction matters. The synchronic wide-scoper says that lacking the means-intention would be (instrumentally) rational, if you lacked the end-intention. But of course, lacking the means-intention would also be rational according to the narrow-scoper, if you didn’t have the end-intention! As Errol Lord (2011) pointed out, the difference is just that while you comply with the wide-scope requirement when you lack the end-intention, the narrow-scope requirement simply doesn’t apply to you when you lack the end-intention. If you lose the end-intention, you escape the narrow requirement. The difference comes out only in diachronic terms, since only the diachronic wide-scoper says that forming the means-intention and giving up the end-intention (or means-belief) are equally (instrumentally) rational moves.³

What about the narrow-scoper? For simplicity, let’s first consider the parallel case of enkrasia. Suppose that on Monday, you form the belief that you ought to travel to Toronto on Wednesday, because you have a meeting there. On Tuesday you learn that it’s cancelled. If a

---

³ This is itself by no means unproblematic, since it means the diachronic wide-scope principle fails to provide any guidance about which attitudes to form or give up. That is why Broome (2013) supplements it with separate basing permissions.
diachronic enkrasia requirement simply says that if you believe at $t_1$ that you ought to F at $t_2$, you are rationally required to intend to F at $t_2$, it entails that you’re still rationally required to intend to travel to Toronto on Wednesday, since you did believe on Monday that you ought to do so on Wednesday. Once $t_1$ has passed, you can’t escape a requirement according to which attitudes at $t_1$ call for attitudes at some later time, since you can no longer change your attitudes at $t_1$. To remedy this, Lord (2013) adds a concurrence clause requiring that one still believes at $t_2$ that one ought to F at $t_2$. This requirement is escapable: since you no longer believe on Wednesday that you ought to go Toronto, rationality no longer requires you to intend to do so. By parallel reasoning, when you no longer intend the end, you’re not required to intend the means either.

However, Benjamin Kiesewetter convincingly argues that a requirement like Lord’s can’t offer rational guidance:

For the relevant condition requires the presence of an attitude state at $[t_2]$; as a result, the requirement does not apply before $[t_2]$. But in order for the requirement to be capable of guiding the relevant adjustments, it must apply at the time the process of adjustment is supposed to begin, not just at the time at which it is supposed to be finished. (Kiesewetter 2013, 193)

I agree. Suppose your meeting isn’t cancelled. It can’t be that it is only Wednesday that you’re required to intend to go to Toronto – if that were the case, you might perfectly rationally have travelled to Thailand on Tuesday, even if that makes it impossible for you to be in Toronto on Wednesday. The lesson is that enkrasia must apply before the time of the action. But how can it then be escapable by a later change of mind? The solution to this dilemma, I believe, is that intentions themselves are cross-temporal commitments. This must be reflected in an accurate formulation of the Action-Intention Principle: when we perceive conclusive reason for a future
action and take it that we won’t perform the action at the right time unless we form an intention
to do so at some prior time, we take ourselves to have conclusive reason to form a future-directed
intentions at the prior time, possibly right now. For example, if I take myself to have reason to go
to shop tomorrow and believe that I won’t go there unless I settle on going there right now and
hold that fixed in my planning for tomorrow, I take myself to have reason to form the intention to
go to the shop tomorrow right now. Given RRPR-D and this understanding of the Action-
Intention Principle, we can derive the following kind of rational requirement (square brackets
added to avoid any ambiguity):

**Diachronic Enkrasia**

If A believes at $t_1$ that she ought to $[F at t_2]$, rationality requires A to form an intention to
$[F at t_2]$ and maintain it at all times between $t_1$ and $t_2$, unless A ceases to believe that she
ought to $[F at t_2]$ at or prior to $t_2$ (and hence changes her mind about reasons for intending
to $[F at t_2]$ prior to $t_3$).

So in our example, if you form the belief on Monday that you ought to go to Toronto on
Wednesday, rationality requires you to form straight away the future-directed intention to go to
Toronto on Wednesday, and hold on to it until Wednesday, unless you change your mind about
what you ought to do. This future-directed intention rules out travelling to Thailand on Tuesday
and other incompatible actions, and requires figuring out how to get to Toronto, so it guides your
actions in the right way. If, however, you find out on Tuesday that the meeting has been
cancelled and thus change your mind about what you ought to do, you are no longer rationally
required to intend to go to Toronto, and are free to book your flight to Thailand.
A proper diachronic formulation of the instrumental requirement will be similar. Strictly speaking, when we intend an end, we’re committed to treating the presumed fact that taking a means at a specific time prior to the time of the end-action (or simultaneously, in case the means are constitutive) facilitates the end-action as a reason to perform the means-action at that time. Bearing in mind the above formulation of the Action-Intention Principle, applying RRPR-D results in the following:

**Diachronic Instrumental Requirement**

If A intends at \( t_1 \) to \( [E \text{ at } t_3] \) and believes at \( t_1 \) that \( [M\text{-ing at } t_2] \) is necessary for her to \( [E \text{ at } t_3] \), and believes that she will only \( [M \text{ at } t_2] \) if she intends at \( t' \) \((\text{where } t_1 < t' < t_2)\) to \( [M \text{ at } t_2] \), rationality requires A to form an intention at \( t' \) to \( [M \text{ at } t_2] \) and maintain it at all times between \( t' \) and \( t_2 \), unless she gives up her intention to E or one of her beliefs prior to \( t_2 \).

(As always, whether it is rational to give up the end-intention depends on your take on reasons for it, which is not a question for instrumental rationality.) This kind of requirement can offer guidance. If you intend to sit on the Iron Throne tomorrow, believe that it is necessary for sitting on the Iron Throne that you sail to Westeros today, and believe that you will only sail to Westeros today if you right now intend to sail to Westeros, rationality requires you to intend right now to sail to Westeros, and hence to rule out alternative plans. At the same time, the requirement is escapable: if, in our original example, you give up the intention to get to work prior to needing to drive over the child, presumably because you realize the costs outweigh the benefits, you’re no longer rationally required to intend to drive over the child. Phew!

4. From Structural to Substantive Rationality: Rationality as Government By Reason
I’ve argued that the perceived reasons view can go further than generally thought, when it is combined with plausible assumptions in the philosophy of mind. But it still has disadvantages that motivate adding the condition of competent perception of reasons.

First, the problems. As a kind of narrow-scope view, RRPR licenses bootstrapping rational requirements into existence. The issue is the most obvious when it comes to capricious beliefs about reasons. Consider a Parfitian case: Elsa has a choice between a mildly inconvenient operation on Monday or a very painful one on Tuesday. She happens to believe she has no reason to avoid pain on future Tuesdays, so she believes the inconvenience of the operation on Monday is a conclusive reason to choose the Tuesday operation. Now, by RRPR, rationality requires her to choose the Tuesday operation. Yet it seems that it would be rational for Elsa not to choose it (indeed, it might be irrational for her to do so). Relatedly, someone who happens to believe she ought to F and not-F is rationally required to intend both F and not-F, which seems clearly irrational (Broome 2013). Second, the perceived reasons account offers a highly deflationary “transparency account” (Kolodny 2005) of the normativity of rationality. In Scanlon’s words, “Normativity enters only from the point of view of the person who has these attitudes, and therefore sees the relevant considerations as reasons.” (2007, 87) To criticize someone for being irrational, on this account, is simply to make a “descriptive, psychological claim” that “from his point of view … he has conclusive reason to have the attitude” (Kolodny 2005, 557), that is, to call his attention to what he himself already thinks. But as Kiesewetter puts it, criticizing someone isn’t just pointing out that she has violated her own standard, but assumes that the standard is authoritative for the target of criticism. I don’t agree with Kiesewetter’s own view that this requires the target to have a decisive reason to do what rationally requires of them, since
I take irrationality to be a distinct kind of failure, but failure to meet one’s own standards doesn’t seem to capture the normative force of the charge of irrationality.

Both of these problems could be avoided if having a proper order in one’s mind required responding correctly to the reasons there objectively are for us. Such reasons do not require us to do crazy things, and are genuinely normative for us. Yet collapsing the distinction between requirements of rationality and requirements of reason comes at too great a cost. Failures of rationality are a distinct sort of normative mistake. For example, it is overwhelmingly plausible that if you have good grounds for believing that eating carrots serves your interests and therefore intend to eat a carrot, you are not irrational, even if unbeknownst to yourself, you have become seriously allergic to carrots overnight, so that there is decisive reason for you to eat something else (Broome 2007). So the link between rationality and correctly responding to objective reasons has to be more indirect.

Various proposals along these lines have been made. Perhaps rationality is a matter of responding correctly to our non-normative beliefs, such that rationality requires us to F just in case if things are as we suppose, there is conclusive reason for us to F (Schroeder 2009, Way 2010, Parfit 2011). But this view classifies us as irrational too easily. In the case above, I’m not necessarily irrational if I intend to eat the carrot, even if I know that what I’m about to eat is a carrot, and it is true that if it is a carrot, there is conclusive reason for me not to eat it. After all, I might not be aware that something’s being a carrot is a reason for me not to eat it, given my unknown allergy (Broome 2013).

So perhaps rationality consists in responding correctly to the reasons we possess (or have), where I possess a reason R to F if and only if R is a reason to F, I know that R obtains, and I treat R as a reason to F (Lord 2013). This view handles the carrot case, but it still seems too restrictive.
Suppose that I know that I have promised Larry to return his book tonight, and treat that as a conclusive reason to return the book to him tonight. It seems I’m rationally at fault if I don’t intend to return the book, even if there is as a matter of fact no conclusive reason for me to return it, and consequently no reason I could possess. For example, it could be that in this case my having promised is not a conclusive reason to return the book – perhaps there’s a significant cost to me involved or perhaps Larry doesn’t want the book, and has forgotten all about the promise. It could also be that unbeknownst to myself, I can’t return the book to Larry. Such inability plausibly functions as a disabler for the reason promising would generally constitute. So even if I don’t possess a conclusive reason to return the book, I can be rationally required to do so, given my (reasonably held) beliefs about reasons and knowledge of non-normative facts.

In short, what we need to do is to strike the right kind of balance between the subjective aspect of rationality as having the attitudes we take ourselves to have reason to have, and the objective aspect of rationality as responsiveness to objective reasons. The perceived reasons account leaves out the latter aspect, and the views just discussed fail to give sufficient importance to the former. I believe that the best way to navigate this thicket begins with the observation that our perceptions of reasons can be better or worse in the sense of resulting from a fallible competence with objective reasons, or failing to do so. This type of view has recently been sketched by Kurt Sylvan (2015). Although he frames his discussion somewhat differently, I believe it is well suited to solve the problems I have been discussing.

In more precise terms, the competence that Sylvan regards as relevant to rationality is the competence to treat a consideration as an objective reason of a relevant kind only if it is, if true, an objective reason of the relevant kind (Sylvan 2015, 599). But what is such a competence? Sylvan’s inspiration comes from Ernest Sosa’s work on virtue epistemology. Sosa (2007) argues
that performances that aim at a goal in general are correct when they hit the relevant target, competent when they manifest a disposition to hit the target, and apt when correct because competent. So competence is a kind of disposition. When we competently take something to be a reason, our reason-taking results from a disposition that tends to hit the target – a disposition to take something to be a reason only when it really is a reason. It can be competent without being correct, however: if circumstances are unfavorable, even a competent archer may fail to hit the target.

As Sylvan (2015) observes, there are many dispositions that amount to the relevant kind of competence. The inductive disposition to take an observed regularity to favor belief in a universal generalization is arguably one example on the theoretical side, as is the disposition to take visual appearances in daylight to constitute evidence for the color of things. In favorable conditions, these dispositions result in correct take on reasons. It is a question for substantive normative theory just which dispositions constitute competence in the practical domain. Presumably the tendency to take future agony for oneself or others to constitute a reason against an action is one such disposition. (I’ll discuss patterns of reasons below.) I don’t have a theory of how to individuate such dispositions – perhaps there are only a small number of higher-level dispositions that subsume the ones I have mentioned. But we should bear in mind that we can be competent reason-recognizers in one domain without being such in all domains. This suggests that the broad notion of competence with reasons really splits into many local competences.

Now, for rhetorical ease, I am going to use the label Reason with a capital R to refer to the set of our competences to treat something as a reason if and only if it is one. I’ll grant straight away that this is a broader conception of Reason than is common. Sometimes people use Reason as the name for an inferential capacity exclusively. But there is also precedent in the Aristotelian
tradition for thinking of Reason as a capacity to recognize reasons in general. In any case, this is just a stipulative convention that makes it easier to speak of competent perception of reasons.

In these terms, my final thesis about rationality is the following:

*Rationality as Government by Reason (RGR)*

Rationality requires S to F if and only if S competently takes herself to have conclusive reason to F. (Optionally: Rationality permits S to F only if S takes herself to have sufficient reason to F in the closest possible circumstance in which she competently considers whether to F.)

So how does RGR help solve the problems of the perceived reasons account while giving our subjective take on reasons its due significance? To begin with, RGR still allows a kind of bootstrapping, since our competent take on reasons gives rise to rational requirements, even when it is mistaken. But since the belief that there is no reason to avoid pain on Future Tuesdays won’t result from the exercise of Reason, RGR doesn’t entail that Elsa is rationally required to choose the painful Tuesday operation. Similarly, if your ought beliefs result from competent recognition of reasons, you typically won’t form beliefs according to which you ought to do things you know are incompatible.

When it comes to the *normativity* of rationality, RGR doesn’t say that we have reason to be rational as such, but simply says that whenever rationality demands some A of us, we competently perceive ourselves to have conclusive reason to A. Does competent perception of reasons give rise to a distinctive kind of normative demand, then? I think it plausibly does. Even if you don’t have reason to respond in the way what your Reason tells you to respond, so that you objectively ought not respond that way, there’s a clear sense in which you are criticizable for not
listening to your Reason. Otherwise, you give the right response to your situation only by a dint of good luck. This is grounds for evaluating you negatively. Moreover, in the standard case, in which you go against the reasons you competently perceive, you go wrong by your own lights, just as Scanlon and Kolodny say. But that’s not all. When we criticize you for irrationality, we don’t just say that you stand self-condemned, so to speak. We also *endorse* the capacity you used to arrive at the take on reasons that you go against. Our criticism is thus *not* just a psychological remark. Even though it still doesn’t follow that you have reason to do what rationality requires – you might accidentally do the right thing – the standard provided by your own Reason is not merely evaluative either. You *should* do your best to do what you have most reason to do, and the way to do that is developing, employing, and abiding by your Reason.

Sometimes, to be sure, we also call someone irrational when they haven’t actually competently formed a stance on reasons and then gone against it. This could be either because you have no stance on reasons, or your stance is incompetent. Elsa, who doesn’t care about agony on future Tuesdays, is a case of the latter. I’ve only observed that her incompetent perception of reasons doesn’t suffice to make her intention to have the operation on Tuesday rationally required. But what if she doesn’t perceive any reason *against* choosing the Tuesday operation? It still seems right to tell her that the choice would be irrational, not just that she has decisive reason against it. My tentative suggestion is that the force of such criticism comes from a counterfactual assumption. We’re saying that were Elsa to exercise her Reason, all else being as equal as possible, she *would* find she has decisive reason against choosing the Tuesday operation, in which case she would be rationally required to give up the intention by RGR. (If, on the other hand, she lacks the competence to recognize such obvious reasons, she might not be an apt target
of rational criticism in this domain in the first place.) So if rational permissibility is tied to such counterfactual responses, as the optional clause in RGR says, her actual intention is irrational.

Does RGR tie rationality too closely to the reasons there are for us in the way that competing reasons-responsiveness views do? No, because even competent responses are fallible. Take the case of returning the book to Larry. The stipulation was that in spite of having promised to do so, I have in fact no (sufficient) reason to do it, given the unusual circumstances. Yet it still seems irrational for me to go against my perception of reasons. RGR agrees with this, since the disposition to think that one has reason to do what one has promised characteristically tracks one’s objective reasons. So it amounts to a competence, even if it misfires in these unfavorable circumstances. Consequently, by RGR, rationality requires me to intend according to my (in this instance false) beliefs about reasons.

Since taking ourselves to have reason is as essential to RGR as it is to RRPR, RGR, too, is a self-governance requirement – it tells us to heed our perceptions of reasons when they are competent. As such, it doesn’t say anything about why we should have consistent intentions, for example, or take the means to our ends. But with a proviso, once we complement it with Normative Constitutivism, we can use it to explain structural rationality in the same way as RRPR. (This is one important respect in which my account goes further than Sylvan’s.) The proviso is that we must make the case that the reason-takings involved in having the attitudes are competent, since according to RGR only competent perceptions of reasons give rise to rational requirements. In other words, the question is whether the disposition embedded in Intention Commitment and the Action-Intention Principle, for example, tends to hit the target – for example, that we (fallibly) take ourselves to have reason to take the means when and only when we do have reason to take the means.
There are two situations here to consider. Focusing on Intention Commitment, either the agent has sufficient reason to intend end E, or she doesn’t. In the favorable case, her intention-embedded perception of a reason non-accidentally tracks a genuine pattern of reasons. Such patterns arise when reasons to do certain things systematically derive from reasons to do other things. For example, as Raz (2005) pointed out, regardless of whether an agent A intends an end E, if there is reason for A to E, and M is both necessary (and facilitative) for E and within the A’s power, there is reason for A to M by way of instrumental transmission. In such a case, the disposition to take oneself to have a reason to M is no doubt a component of competence with reasons. Tracking patterns of reasons is a kind of skill that is partially constitutive of Reason. After all, without such dispositions, we would fail to perceive many of the reasons there are for us, and consequently fail to do what we have reason to do, even if we’re rational and correctly perceive reasons for ends.

How about the unfavorable case? There’s a prima facie challenge here, since we clearly can have intentions to do things we shouldn’t do, and correspondingly take ourselves to have reason to take means when we don’t have one. But it doesn’t yet follow that the reason-taking isn’t competent, since in unfavorable circumstances, even competent perceptions will be systematically mistaken – even a skilled archer will systematically miss if she tries to shoot in stormy conditions. (They may also occasionally miss in good conditions, of course.) When it comes to perceptions of derivative reasons, circumstances are unfavorable when we lack a correct perception of the source reason. To be sure, in this case it may be our own shortcoming that is responsible for the circumstances being unfavorable. One way to think about this is suggested by Sosa’s (2011) distinction between apt and meta-apt performances. A performance is meta-apt when it is informed by awareness of the circumstances being favourable. Using his favourite
archery example, he notes that when an archer’s performance reflects her meta-competence for target and shot selection, “it is no accident that the shot is made in specific conditions where the archer’s competence is up to the task of producing success with a high enough percentage” (Sosa 2011, 10). Using this distinction, we might say that our perceptions of instrumental reasons can be competent even when we lack reason for the end, but this competence can be expected to misfire when we haven’t deployed our Reason in selecting the end itself. Here selecting the end involves exercising a meta-competence relative to taking the means.

5. Conclusion

In this paper, I’ve defended the straightforward view that at least an important part of practical rationality consists in having attitudes our Reason tells or permits us to have. (Extending the thesis to theoretical rationality is not a trivial task, and I have not attempted it here.) I’ve claimed that with independently justified assumptions in the philosophy of mind, this view accounts for familiar structural features of rationality, such as the irrationality of incoherent combinations of attitudes, without assuming independent coherence requirements. Insofar as we think of Reason as fallible sensitivity to reasons and patterns of reasons, the account will also explain the intuition that capricious perceptions of reasons do not generate rational requirements.

This intuition, to be sure, is not universally shared. Some may reject the assessment of Elsa as irrational. This will have the additional cost of leaving it obscure why you should do the thing you think you have most reason to do, since it will not even remotely reliably track what there actually is most reason for you to do. Nevertheless, error theorists about normativity and others who are willing to accept these costs may read the first three sections of this paper as a
new defense of the weaker claim that rationality consists in forming attitudes we take ourselves to have sufficient or conclusive reasons to have.

The view of rationality as the rule of reason defended in the final section introduces a substantive element, so that a proper order in one’s mind isn’t merely a matter of fitting second-order and first-order attitudes together, or coherence among first-order attitudes. Nevertheless, it allows for the demands of rationality come apart from what there is actually reason for us to do, and thus respects the distinction between rationality and conformity to reasons as separate dimensions of normative assessment. This, I contend, strikes the right balance between the subjective and objective aspects of rationality.

References


Kim, Jaegwon. What Is


Raz, Joseph 2005.


