UNITY OF REASONS
Adam Cureton

[Draft – December 31, 2012]

Abstract: Reasons and rationality are basic normative notions. Substantive realism about reasons (SRR) explains rationality in terms of reasons while Kantian constructivism about reasons (KCR) explains reasons in terms of rationality. This is a fundamental normative disagreement that is separate from the metaethical discussions of these views that have been prominent. By supposing that reasons are here, there and everywhere, SRR seems to lack sufficient normative unity and structure, apparently ending normative inquiry too early. But KCR seems to face a dilemma of its own: Either a putative conception of rationality is thick enough to capture the reasons of commonsense, in which case it is not a conception of rationality at all, or a putative conception of rationality is genuine, in which case it is too thin to generate the reasons we recognize in commonsense. My aim is to argue for a version of KCR that is built on an expanded conception of rational agency. I develop two complementary strategies to justify incorporating non-formal features into a conception of rationality. The argument from modest constitutivism holds that certain substantive commitments are constitutive of being a rational agent but not necessarily constitutive of every action. The second approach is to lay out a substantive conception of rationality as a normative starting point and begin building a normative theory that may achieve reflective equilibrium. I then illustrate how a substantive account of rationality can generate a wide variety of the reasons that exist with a contractualist procedure that is different from Scanlon’s.

There are at least two basic normative notions: rationality and reasons. The structure of a normative theory depends on how these ideas are interpreted and how they are related to each other.¹ Some normative theories explain rationality in terms of reasons while others explain reasons in terms of rationality; some say an agent is rational because she responds correctly to (believed) reasons while others say that facts are made into reasons by rational agents.

The dominant account of reasons nowadays, which is known as substantive realism about reasons (SRR), holds that reasons are normatively basic, they come in many shapes and sizes, and there is no underlying normative explanation of them in terms of other normative notions.² Rationality, on this view, is a matter of correctly responding to reasons that exist or that one takes to exist. As T. M. Scanlon puts it:

---

¹ This way of describing the distinctive aims and subject-matter of normative theory generalized from Rawls’ understanding of moral theory: “There are three basic moral notions: the good, the right, and the morally good (moral worth of persons). The structure of a moral theory depends on how these notions are defined and related to one another. It is characteristic of teleological theories that they start with an independent definition of the good and then define the right as maximizing this good.” (Rawls 1999a: 242)

² Proponents of substantive realism about reasons include (Dancy, 2004; Nagel, 1978; Parfit, 2009; Raz, 1999; Scanlon, 1998)
It seems to me that there always remain substantive normative questions about ‘What is a reason for what?’ which must be faced, and answered, directly. This is why I am not a Kantian, or a neo-Kantian of Korsgaard's sort. (Scanlon, 2003a, p. 15).

Kantian constructivism about reasons (KCR), is the reverse view: Rationality is the primitive normative notion that picks out which facts are reasons for what and explains why those normative relations hold.³ As far back as 1951, in his review of Stephen Toulmin’s book *The Place of Reason in Ethics*, John Rawls endorses a version of KCR:

> The point is that *a reason is any consideration which competent persons in their reflective moments feel bound to give some weight to whether or not they think the consideration sufficient in itself to settle the case* (Rawls, 1951, p. 577, his italics).⁴

How are we to decide between substantive realism about reasons and Kantian constructivism about reasons? I think there is a fundamental normative disagreement between them that is separate from the metaethical discussions that have been prominent in the recent literature.⁵ Just as the dispute about the normative priority of the right and the good cannot be settled on metaphysical or epistemological grounds⁶, so must we engage in normative theorizing to decide the normative priority of rationality and reasons.

SRR and KCR face dueling normative challenges that pull in opposing directions. According to the *unity objection*, by assuming that reasons are here, there and everywhere, SRR lacks sufficient normative unity and structure. It ends normative inquiry too early, closing off the possibility of an underlying normative explanations of why, for example, the fact that someone is my spouse is a reason to be attentive to her needs, why the fact that smoking kills is a reason to quit or perhaps even why the fact that doing something would cause pain is a reason not to do it.⁷ On the other hand, according to the *particularity objection*, proposals for reducing reasons to other normative ideas such as rationality are too limited and anemic to capture the wide variety of reasons that are recognized in commonsense. Specifically, as Scanlon says against KCR: “I do not believe the idea of rational agency is rich enough to yield all the claims about reasons that seem evidently correct” (c.f. Scanlon, 2003a; Scanlon, 2003b, 2007; 2011, p. 124).

The aim of this paper is to argue for a version of Kantian constructivism about reasons that is built on a substantive conception of rational agency.⁸ By expanding the notion of

---

³ (Hill, 2012a; Korsgaard, 1996c; Rawls, 1999f; Street, 2008). I mostly set aside theoretical reasons, although a fully worked-out Kantian constructivist view would have to show how those reasons are also explained by rationality.

⁴ This same idea is implicit in (Rawls, 1999d) and explicit in (Rawls, 1999c).

⁵ This has been a theme in Scanlon’s work that I endorse. See especially (Scanlon, 1998, 2009). For meta-normative discussions of SRR and KCR, see (Darwall, Gibbard, & Railton, 1992; Enoch, 2011; FitzPatrick, 2005; Lavin, 2004).

⁶ (Rawls, 1999a, p. 286; 1999c, p. 343; 1999e, p. 242). Simple consequentialism, for example, specifies a conception of value that is independent of what is right and holds that right acts are those that maximize the good whereas simple deontology reverses this order of explanation by assuming that right acts are defined independently of the value of states of affairs and claims instead that what counts as good is constrained by principles of right.

⁷ Because of the apparent interconnected nature of various basic normative concepts, it may be that the unity objection cannot be evaluated in isolation from other aspects of the normative theories in which it figures, so on reflection it may be best to abandon the search for greater unity if there are sufficient payoffs in other areas for doing so.

⁸ The most common conceptions of rationality in the contemporary literature make rationality purely formal, a matter of the consistency and coherence of our mental states (Broome, 2010; Kolodny, 2005; Smith, 2004) although there is a Kantian tradition that makes rationality a matter of being reasonable as well (Hill, 1991, 1992a, 1992b; Rawls, 1999c, 1999f).
rationality to include non-formal features, KCR stands a better chance at unifying the particular reasons we would endorse on due reflection. More specifically, I begin by characterizing how SRR has, again and again, successfully resisted attempts at reducing reasons to different normative notions, such as values, desires or principles. Then I deploy their basic strategy against KCR in the form of a dilemma: Either a putative conception of rationality is thick enough to capture the reasons of commonsense, in which case it is not actually a conception of rationality at all, or a putative conception of rationality is genuine, in which case it is too thin to generate the reasons we recognize in commonsense. I argue that two prominent versions of KCR, those of Onora O’Neill and Rawls, fall on one horn of this dilemma or the other while Korsgaard’s constitutivist view, in which a commitment to unconditional rational standards is constitutive of agency, is promising but depends on a questionable account of action that excludes the possibility of knowingly and willingly doing wrong. The solution to the KCR dilemma, I claim, is to expand our notion of rationality to include substantive as well as formal elements. Restricting my scope to just three features, self-preservation, exercise of one’s rational capacities and reciprocity, I develop two complementary strategies for incorporating these elements into a conception of rationality. The first argument claims that a commitment to self-preservation, to exercising one’s rational capacities and to reciprocating with others are constitutive of being an autonomous agent rather than constitutive of each and every action.\(^9\) The second approach is to follow the lead set by defenders of SRR by laying out a plausible starting point in normative theory, explaining how it fits with commonsense, and showing that a normative theory built on that basis is more likely than others to achieve full and wide reflective equilibrium. I conclude by illustrating how a conception of rationality that includes self-preservation, exercise of one’s rational powers, and reciprocity can generate a wide variety of reasons of commonsense through a moral contractualist procedure that is different from Scanlon’s. **Substantive realism about reasons**

Defenders of SRR have a powerful two-part strategy against attempts at reducing reasons to values, desires or principles about what we ought to do: First, there is the particularity objection, which says that the reasons that exist are just too complicated to be explained by these other normative notions. Second, the tables are then turned by suggesting that these concepts are actually better explained in terms of reasons rather than the other way around.\(^10\) According to the SRR strategy, the reasons of everyday life are too varied, complicated and specific to be captured by other normative notions, so we must take reflective commonsense at its face and admit a plurality of unexplained reasons, but fortunately treating the notion of a reason as primitive proves quite fruitful in understanding various normative notions.

One might start, for example, with the idea that states of affairs are intrinsically valuable and then try to explain all the practical reasons there are in terms of promoting better ways the world might go. The first prong of the SRR strategy is that many of the reasons we recognize resist this type of teleological explanation, such as those of friendship; our reasons also seem to come from a variety of sources, which makes it unlikely that they all stem from a single value property such as human welfare. The second prong is a buck-passing counter-proposal, which starts with a wide variety of reasons we have to adopt positive attitudes towards things and to

---

\(^9\) Discussions of constitutivism have mostly focused on constitutive aspects of action whereas I am interested in constitutive aspects of being an agent of a certain kind: See (Enoch, 2006; Ferrero, 2009; Rosati, 2003; Setiya, 2003)

\(^10\) The SRR strategy is most prominent in (Scanlon, 1998).
treat them in certain ways, and then explains the value of something in terms of the reasons its natural features give us to value it rather than the other way around.\textsuperscript{11}

Another possibility is that when someone has a reason to do something, this is because doing so will help to fulfill her desires. According to the first part of the SRR strategy, there are a variety of reasons we have that do not depend on our having desires, such as ones to preserve a natural landscape or help one another, and we sometimes find ourselves with desires that we evidently have no reasons to satisfy, such as a desire to drink the clear liquid that is actually gasoline. The second prong is that, instead of explaining reasons in terms of desires, it is better to say that desiring something, in at least one sense, is for one’s attention to be directed towards apparent reasons in its favor.

The SRR strategy also guides Scanlon’s contractualist moral theory and reveals the essential role that SRR plays in his normative framework. An opponent of SRR might begin with a set of principles about how we ought to act and then define reasons as facts that are picked out by those principles – we ought to help other people, for instance, so the fact that this person is in need is a reason to provide assistance.\textsuperscript{12} The first part of the SRR strategy is that the correct set of normative principles is too anemic to capture all of the reasons there are: We have various reasons to value things that cannot be fully captured by moral or prudential principles about what we ought to do; we seem to have reasons to act in ways that are nonetheless forbidden all things considered; and perhaps no set of genuine principles could codify all of the reasons that exist because those principles would be so riddled with exceptions as to become meaningless.\textsuperscript{13} The second prong is then to argue that what we ought to do is actually better explained in terms of reasons rather than the other way around. One natural suggestion along these lines, which some defenders of SRR have endorsed, is simply that what we ought to do is what we have most reason to do.\textsuperscript{14} Scanlon’s contractualist account is more complicated than this, but according to him what we owe to each other is still a function of independently specified, free-floating reasons. He starts with a limited set of reasons that particular individuals have in virtue of their various generally described points of view. Included in this privileged group of basic, unexplained reasons are self-regarding ones, such as to promote one’s own wellbeing and exercise one’s own freedom, along with other-regarding ones, such as to treat others fairly; excluded from the set are reasons that do not have to do with the interests of particular individuals, such as those arising from animals or groups of people. When we consider candidate principles about how we may act, we are supposed to note the various generic reasons that individuals have for or against them and then compare these reasons to determine if it is reasonable to reject the principles or not. The notions of right and wrong, according to Scanlon, are explained in terms of principles that are justifiable to everyone on the basis of this class of reasons.\textsuperscript{15}

The best case for SRR is therefore a holistic one: SRR is unwilling to sacrifice the complexity and variety of reasons we recognize in commonsense and that view is deeply embedded in compelling accounts of other normative notions. The evident plausibility of a comprehensive and interconnected normative theory that treats reasons as primitive provides

\textsuperscript{11} Scanlon calls this a ‘buck-passing’ account of value.
\textsuperscript{12} (Hill, 2012a)
\textsuperscript{13} (Dancy, 2004)
\textsuperscript{14} Those who have endorsed a version of this view include (Dancy, 2004; Parfit, 2001; Raz, 1999).
\textsuperscript{15} These reasons and their relations to one another help to define and give determinate content to the abstract idea of acting only in ways that are justifiable to others.
strong indirect support for SRR itself. Even if a distinct normative notion could somehow capture the reasons of commonsense, strong reasons remain for thinking the order of explanation would still begin with reasons. On the whole, according to defenders of SRR, normative theories that admits reasons as starting points are more likely than any others to match our considered normative judgments on due reflection.

**Kantian constructivism about reasons**

There may still be doubts that SRR relies too heavily on intuitions about particular reasons and specific judgments about how reasons combine together, arbitrarily cutting off normative ‘why’ questions that could help to unify and structure the reasons that exist. W.D. Ross argued for a version of SRR that seems sensible as a sort of default position in ethical theory, but there is a natural drive for an underlying explanation of the “unconnected heap” of *prima facie* duties he proposed and for more determinate guidance about what to do when they conflict. 16

How should we assess SRR if we cannot easily disentangle it from the core of various reasonable normative theories? Until we have studied competing normative theories that attempt to avoid the unity objection and the particularity objection, we cannot be sure whether SRR is correct or not. Such competing normative theories, if they are possible, aim to impose more structure on reasons while attempting to capture the complexity and variety of reasons we recognize in commonsense. The best offense against SRR is a good defense of a competing view.

Kantian normative theories offer a reasonably well-developed, but very different, alternative to ones that incorporate SRR. Kantian constructivism about reasons (KCR), which is an essential element of such views, starts with *rational agency* as the most basic normative notion and uses that idea to explain how fully rational persons, on their own or together, could or would reason about and render judgment on normative issues under certain specified conditions. Reasons are facts that figure in their deliberations or are picked out by the principles they specify or the judgments they make when deliberating on the basis of their rational nature and the facts of the case. Appropriately responding to independent reasons is not what makes someone rational; instead, the constitutive rational principles, deliberations and judgments of rational agents provide a unified framework that determines which facts are reasons for what.

Defenders of SRR have a strong case against attempts at reducing reasons to values, desires and principles about what we ought to do, but why are they also skeptical of explaining the reasons that evidently exist by appeal to the idea of *rational agency*? Perhaps their concerns can be partially captured in the form of a dilemma for KCR:

(1) If a putative conception of rationality is genuine then it will be too thin to capture the reasons of commonsense.

(2) If a putative conception of rationality is robust enough to generate all the reasons there are then it is not actually a conception of rationality at all.

(3) Therefore, no genuine conception of rationality can explain all of the reasons that seem evidently correct.

Two of the most prominent versions of Kantian constructivism about reasons fall on separate horns of this dilemma.

(1) Onora O’Neill’s (1990; 1996; 2003a; 2003b) version of Kantian constructivism emphasizes that Kant’s critical method is fundamentally a rejection of arbitrary authority,

---

16 Moral particularists such as Jonathan Dancy (2004) go a step further by claiming that there are a wide variety of irreducible reasons that can even switch their valence depending on the circumstances.
whether that of God, other people, our own feelings, or ‘intrinsic’ values and reasons that exist independently of us. The use of reason, she argues, must be ‘vindicated’ on its own terms through an on-going, public process whereas other proposed sources of normativity cannot justify their own authority in a way that is acceptable to reason. The only normative starting point that survives reason’s critique is the supreme principle of reason itself, which O’Neill takes to be a very thin and quasi-formal version of the Formula of Universal Law, to act only on maxims that we can at the same time will that they be universal laws. More specific moral principles are constructed by reasoning on the basis of this principle, which requires us to test whether our maxims can be consistently shared and rationally communicated to everyone in the actual world. The constructive procedure O’Neill describes is sparse because it does not include idealizations such as a ‘veil of ignorance’, a ‘well-ordered’ society or appeals to what hypothetical agents ‘would’ choose as opposed to what any rational person ‘could’ choose.17

In response to the first side of the dilemma, O’Neill can plausibly argue that the Formula of Universal Law, in the relatively thin form she endorses, is a constitutive principle of rationality. The principle is not altogether formal, but it is close, by requiring us to will only maxims that, if they were made into universal laws, could possibly be consistent with each person’s actual set of maxims and with the standards of rational willing, which are themselves more or less concerned with the consistency and coherence of each agent’s own mental states.18

Once we confine ourselves to a quasi-formal conception of rationality, however, the second part of the dilemma looms, for we must now explain how a procedure grounded in rationality of this kind is thick enough to generate the rich diversity of reasons we recognize in commonsense.19 This is the old charge of empty formalism – how can such a seemingly sparse and anemic principle as the Formula of Universal Law capture and explain all of the particular requirements of morality, let alone all of the reasons that we recognize in commonsense? For one thing, O’Neill’s version of FUL only rules out maxims, it does not seem to tell us what positive actions we ought to take (O’Neill, 1990, p. 103). Identifying an agent’s maxim is notoriously difficult and maxims can often be easily redescribed in ways that threaten to rig the results of the procedure. There are also various “false-positive” examples, such as “I will commit violent acts for revenge”, and “false-negative” ones, such as “I will play tennis on Sunday mornings to avoid crowds.”20 O’Neill has been admirably working to answer the concern raised by the second horn, but many of us remain skeptical about the prospects of explaining the reasons there are on the basis of such a limited conception of rationality.

(2) Rawls’ proposes a version of Kantian constructivism that looks to common moral consciousness and our public political culture to specify a conception of rational agency that is far more substantive and moralized than O’Neill’s. According to him, we have two senses of ‘rational’ in ordinary language, a formal one that means logical and acting for one’s own good and a more general one that means ‘based on reason’ and so includes formal rationality along with reasonableness as well. Being reasonable involves being sane, judicious, fair-minded and sensible. Rawls illustrates the distinction between formal rationality and reasonableness with the

17 For example, no one can possibly consent to being deceived, according to O’Neill, so there is a presumptive prohibition against maxims that incorporate deception.
18 We are rationally required to will the necessary means to our ends or give them up, according to O’Neill; she once thought we must will the normal and foreseeable results of our actions; and her current supplement is that if we are rational we implement our fundamental maxims in ways that are consistent with itself, our specific intentions and foreseeable results.
19 This kind of objection is raised by (James, 2007; Lavin, 2004; Setiya, 2003)
20 (Herman, 1993; Kerstein, 2002)
example of driving a hard bargain, which can be quite rational in the formal sense because it is in one’s interest, though it is sometimes unreasonable, because it is unfair to take advantage of someone’s inferior bargaining position.

Rawls’ conception of rational agency is quite thick because a rational agent, according to him, is both rational in the formal sense and reasonable as well. More specifically: (A) Reasonable people are “are ready to propose, or to acknowledge when proposed by others, the principles needed to specify what can be seen by all as fair terms of cooperation” (Rawls, 2001, p. 191). Rational people, on the other hand, have their own ends, loyalties and attachments, but when they are reasonable as well they are willing sometimes to sacrifice their own interests in order to honor fair terms of cooperation as long as others are willing to do the same.21 (B) A reasonable person is willing to listen to reasons offered by others as such and to recognize ways in which reasonable people may disagree with one another, whereas a purely rational person is only willing to do so if this promotes his interests. (C) A reasonable person has a regulative desire and higher-order interest to fully realize her rational and reasonable nature by developing and exercising all of her powers of reason, which involve forming and pursuing a conception of the good within the constraints of mutually recognized fair social cooperation in which she is seen as an active participant. And (D) a reasonable person develops more and more trust and confidence in social arrangements that they and others willingly accept, comply with and regard as fair and just.22 What it is to be a reason in general, Rawls might say, is to be explained by the deliberations, judgments or choices of reasonable and rational agents in virtue of their nature as such.23

By specifying a substantive conception of rational agency, one that includes various unconditional and non-formal commitments, dispositions, motivations and interests, Rawls may have enough materials to capture the reasons of commonsense with the help of one or more constructive procedures. But if he can avoid the second part of the dilemma in this way then the first part of the dilemma becomes more daunting – why should we think that this thick idea is a conception of rationality at all? Any attempt to justify such a significant expansion of rationality, it seems, will be arbitrary, circular or render KCR uninteresting.24 For one thing, if Rawls’ conception of the person is taken from the public political culture of the historical period in which we happen to live, as he sometimes suggests, then his putative conception of rational agency seems to be arbitrary, contingent and lacking in normativity. Second, if the proposed supplements are adopted because we recognize that there are reasons for us to treat ourselves and one another in certain ways then the resulting conception of rational agency would be circular if we later claim that it is the source of all reasons. Third, if we make the idea of rational agency too thick by incorporating most all aspects of normativity into a proposed account of rational agency – perhaps by defining an ideal of a fully virtuous person with perfect moral judgment – then KCR would be far less interesting as an explanatory claim about the nature of reasons.

21 In this way, according to Rawls, the reasonable subordinates the rational.
22 Rawls calls this a “reasonable moral psychology” (Rawls, 1993, p. 82) which he develops for the specific problem of justice but can also be extended to matters of interpersonal morality as well.
23 The main features of Rawls’ original position are described as structural representations of these features of the person.
24 Objections of this sort are raised by (Enoch, 2006; FitzPatrick, 2005; James, 2007)
One way to avoid the dilemma for Kantian constructivism about reasons may be Christine Korsgaard’s constitutivist strategy, but her view comes at the expense of a questionable theory of action.\textsuperscript{25}

According to Korsgaard (2009a, 2009b), the Categorical Imperative, as the supreme principle of rationality, is constitutive of all action, so our actions are necessarily guided by what is most objectively and subjectively good and reasonable. Korsgaard argues roughly as follows: When we deliberate, we suppose that we can stand apart from our various incentives to act, and when we make a choice, we identify ourselves with the principle on which we chose. If we choose to act on a particular incentive that we take as our own reason only in this specific case, without thereby counting it as a universal reason that holds for us and others in other cases as well, then there can be no difference between us and the incentives on which we act. This would mean we would not be determining ourselves to act at all, so we would have no will if we only choose on the basis of personal reasons. When we exercise our freedom, therefore, we necessarily will only universal laws, which we thereby identify ourselves with. But, because our wills are free, the universal laws that we identify with cannot be imposed by anything outside of ourselves, so the will must be its own law, it must be autonomous. And the only principle that can be the most basic principle of an autonomous will is FUL – “a free will and a will under moral laws” says Kant “are one and the same” (G 4:447).

When we act, we always take ourselves to be acting on maxims that can be willed as universal laws, but this does not mean that what we do is always best. In some cases, we may be swept away by natural impulses, which would mean we did not exercise our will in behaving in that way, for the will is identified with practical reason. And there can be misuses of our practical reason when we think our maxims are morally justified even though they are not, in which case our choices are not genuine acts of will but rather defective actions that were aiming to be morally appropriate.

On Korsgaard’s account, reasons are constructed in the following way:
To put it in my own terms, when an agent determines whether she can will a maxim as universal law, she is determining that she can endorse a certain consideration in favor of doing something and therefore can treat it as a reason (Korsgaard, 2003, p. 118)

There are various questions we might have with Korsgaard’s challenging argument, but if her version of constitutivism can be justified then it offers a compelling way of resolving the dilemma for KCR. On the second horn, if Korsgaard is correct that whenever we act we thereby take our reasons to be reasons for everyone, and so necessarily regard our maxims as universal laws, then her view justifies an expanded notion of rationality by explaining how that way of understanding rationality is constitutive of action itself. On the first horn, Korsgaard has drawn out several implications of her thick conception of rationality that, according to her, suggest how it may be able to generate all of the reasons that exist. She has argued, for example, that there are standards of rational willing that govern the purposes that underlie our actions, so her version of Formula of Universal Law is more robust than O’Neill’s because it forbids a wide variety of maxims that would, if they were universalized, be self-defeating. Korsgaard has also described a ‘regress argument’ for the Formula of Humanity by claiming that acting for universal reasons commits us to regarding rational nature in general as the source of objective value and so as objectively valuable itself.\textsuperscript{26}

\textsuperscript{25} (Korsgaard, 1996c, 2009a, 2009b)
\textsuperscript{26} See (Korsgaard, 1996a, 1996b)
Korsgaard’s brand of constitutivism may be a promising strategy for avoiding the KCR dilemma, but it comes at the expense of a conception of action that some may find implausible. As Thomas Hill has argued, Korsgaard’s view does not allow for cases of weakness of will in which we knowingly and willingly do wrong. It seems evidently possible that someone who is committed to morality may nonetheless knowingly decide to act contrary to it, and even do so fervently, with gusto and resolve. It is difficult to see how all acts of deliberate cruelty, for example, are actually failed attempts at doing the right thing.

In sum, the prospects for KCR look bleak. O’Neill may have described a genuine conception of rationality but it is probably too thin to capture the reasons of commonsense; Rawls’ idea of the person may be thick enough to explain those reasons but he may not have actually given an account of rationality; and Korsgaard may have successfully described a conception of rationality that is also sufficiently robust but at the expense of excluding certain kinds of acts, such as those of deliberate cruelty, from the space of reasons altogether.

A Kantian argument for substantive rationality

Defenders of KCR, it seems, must radically pare down which considerations are in fact reasons, or they can try to justify expanding the notion of rationality in some other way. The first option is unacceptable, but if rationality did include non-formal elements, those that are not just about consistency and coherence among mental states, then there would be some hope that rationality could capture the reasons we recognize in commonsense.

How else might one argue that there are substantive features of rationality? Let’s focus on three putative ones:

- **Self-preservation** – Rationality presumptively requires an agent to be committed to preserving its own existence as a rational agent.
- **Exercise** – Rationality presumptively requires an agent to be committed to exercising its rational capacities as a rational agent.
- **Reciprocity** – Rationality requires an agent to be committed to acting only in ways that are justifiable to other rational agents.

My aim in this section and the next is to carve out a toe-hold for these three specific features of rationality as a way of responding to the second horn of the dilemma for KCR; in the final section I return to the first horn of that dilemma by illustrating how a conception of rationality that incorporates them can generate a wide variety of the reasons we recognize in commonsense.

Here are a few points of clarification about how I understand self-preservation, exercise and reciprocity. First, the notion of rationality I am explicating is the broad one Rawls describes of ‘having to do with reason’, which includes reasonableness along with rationality in the economists’ sense. Second, these requirements are not formal, they are not about good mental house-keeping, but are instead substantive (putative) requirements to develop and maintain unconditional commitments. Third, having a commitment in the relevant sense is a matter of setting a policy for oneself and being disposed to act accordingly, even though an agent may not always choose to follow her own standards. Fourth, reciprocity is an all-things-considered requirement that takes precedence over self-preservation and exercise, which are presumptive requirements that can be defeated in various contexts by other features of rationality. I have not

27 (Hill, 2012b)
28 Aaron James (2007) favors abandoning purely formal conceptions of rationality as well, and the substantive elements he argues for (including robust notions of judgment, attention and open-mindedness) seem quite plausible to me as further non-formal features of rationality in addition to the ones I argue for here.
specified the relative priority of these defeasible commitments, so it may be rational, for example, to endure great personal risk if doing so is required by other requirements of rationality. Finally, self-preservation and exercise are self-regarding requirement of rationality that generate a rational interest on an agent’s behalf, but they do not as such imply that a rational agent must be concerned to preserve other rational agents or help them to exercise their rational capacities.

There are two complementary ways to argue that self-preservation, exercise and reciprocity are genuine features of rationality. The first is a straightforward argument that relies on thoroughly Kantian premises, while the second is an indirect and holistic approach to structuring and systematizing considered normative judgments around the commonsense idea of a reasonable and rational person. The strategies are related because, as I see it, there is a Kantian interpretation of our ordinary notion of a reason-governed person, but one need not be a Kantian to take that commonsense idea as a starting point in normative theory.29

First the Kantian argument from modest constitutivism. Constitutivists try to justify substantive conceptions of rationality by arguing that freedom necessarily involves having non-formal commitments and dispositions of rationality.30 Kantians defend this claim by appeal to the vexing argument Kant gives at the beginning of *Groundwork* 3.31 There are different ways of understanding his line of reasoning, however, and one of them suggests a distinct, and in some ways more appealing, type of constitutivism than Korsgaard’s. According to modest constitutivism, substantive rational commitments and dispositions are not constitutive of each and every action, as Korsgaard thinks, but they are constitutive of being an agent.

The argument for modest constitutivism can be stated simply, although the premises will require some explanation:

- (1) Rational agents necessarily have autonomy of the will.
- (2) Self-preservation, exercise and reciprocity are necessary commitments of an autonomous agent.
- (3) Therefore, they are constitutive features of rationality as well.

Consider premise (1), which asserts autonomy as the middle term between rationality and substantive requirements such as self-preservation, exercise and reciprocity. There is broad agreement among Kant’s commentators on how Kant attempts to defend this claim: A rational agent who engages in any practical thought whatsoever thereby takes up a first-person, deliberative perspective where she regards herself as having a free will, which is a capacity to make things happen in the world “independently of alien causes determining it” (G 4:445). This kind of negative freedom is inseparable from “positive freedom” or autonomy of the will because the choices of a rational agent cannot be lawless, but nor can they be determined by natural inclinations and sympathies, selfish desires, God’s commands or external moral facts. The choices of a rational agent must therefore be explained by laws of a different kind, namely principles of pure practical reason. Thus, rational agents necessarily have autonomy of the will, understood as “the property the will has of being a law to itself (independently of any property of the objects of volition)” (G 4: 440). And the supreme principle of a rational, autonomous will,

---

29 Rawls (1999f, pp. 221-227) offers a Kantian interpretation of his conception of justice as fairness, which he thinks can be established on independent grounds as well.
30 David Velleman (2000); (Velleman, 2007, 2009) may endorse a constitutivist view as well, although it is unclear whether his view is Kantian enough to vindicate KCR.
31 (Kant, Hill, & Zweig, 2002), references will be given as G followed by the Academy volume and page number.
which must be a necessary, unconditional and universally binding rational norm, is the Formula of Universal Law (G 4:419-421).

Constitutivists such as Korsgaard accept a stronger interpretation of this argument, which they think establishes that when we perform any action whatsoever, we are unfailingly directed to what is most rational and reasonable. As they see it, our wills cannot be lawless, determined by inclination or directed by God, so whenever we act freely, we are acting in accord with principles of pure practical reason – a free will, according to them, just is practical reason. When we are swept away by our emotions, we are not free, but when we act wrongly, we are merely misusing our rational capacity by “choosing” (but never “willing”) to do things that we implicitly but incorrectly take to be objectively good. What is not possible, on this account, is to willingly act in ways we know to be immoral.

A weaker reading of Kant’s argument is also possible, however, once we distinguish between the will of a rational agent, understood as a set of commitments that an agent accepts along with dispositions to follow them, and particular acts of will, in which we choose whether or not to follow our own basic standards. On this construal, Kant’s argument is about the constitutive features of an agent’s will, not constitutive features of specific acts of will. In order to have a will at all, we must have basic policies that we take to govern our particular acts of will, but the accepted laws of our will cannot be determined by “alien” causes, so they must be self-legislated laws of practical reason that we give to ourselves. In other words, a free will is necessarily an autonomous will. But even though a rational, autonomous agent necessarily imposes moral requirements on herself, she also has the ability to temporarily ignore them or set them aside. She can know that she is acting selfishly, irresponsibly or reprehensibly without aiming at what is right and good in those particular acts of will – as a rational agent, she inevitably sets moral policies for herself that she accepts but can on occasion flaunt. When she chooses to violate her most fundamental commitments, she is internally conflicted and usually left feeling regret, guilt and pangs of conscience. She is most fully free, however, when she expresses her autonomy in action by following the moral laws that she gives to herself.

According to modest constitutivism, agents are constitutively autonomous, which means they are fundamentally committed to the laws of morality as self-imposed rational standards, but rational agents can sometimes knowingly and temporarily stray from the basic moral commitments that they unavoidably will for themselves. Unlike Korsgaard’s view, we are not aiming at what is reasonable and rational in each and every act, but in being a rational agent we are constitutively committed to self-imposed moral standards that we legislate for ourselves even though we sometimes knowingly and willingly ignore them. Modest constitutivism fits our commonsense views about action more closely than Korsgaard’s strategy, but it still asserts a constitutive connection between rational agency and morality that defenders of KCR can use to explain how the notion of rational agency is thick enough to generate the reasons that seem evidently correct.

Now consider premise (2), which says that autonomous agents are by their nature committed to preserving themselves, exercising their rational capacities and reciprocating with one another then.

Take self-preservation. An autonomous agent not only has the capacity to be a law unto herself, she is also disposed, in virtue of her nature as such, to select certain laws that she regards as authoritative and binding on her actions and to act in ways that conform to her own acknowledged standards even though she might not do so. Her continued existence as an autonomous agent is often a necessary condition for fulfilling her disposition to be self-
legislating, so she is also committed to preserving herself in such cases. Part of being an autonomous agent is to have a desire to be self-governing, to resist constraint from external authorities, so autonomous agents take an interest in their own continued existence as a necessary part of achieving self-mastery. If she allows her freedom to deteriorate or waste away, she becomes part of the causal nexus rather than governed by self-imposed rational standards.

When it comes to exercising one’s rational capacities, an autonomous person is most fully free when she chooses to follow the basic moral standards that she legislates for herself, so she is disposed to protect, develop and exercise the capacities that make her the author and subject of the moral law. An autonomous person is also disposed to think for herself, to settle on her own non-personal plans, projects and commitments, and organize and schedule her ends, so she takes an interest in exercising these capacities as well. Self-preservation and exercise are defeasible features of autonomy that may be overridden, for example, when reciprocal moral requirements require an agent to sacrifice herself or when an agent’s ability to function as a rational, autonomous agent is degenerating quickly.

Autonomous agents are necessarily committed to reciprocating with others. The supreme principle of an autonomous will is supposed to be a necessary, universal and unconditional rational standard that is not imposed by an external authority. All autonomous wills are in this same position – they are indistinguishable as such except that they have their own self-regarding interests. So the most basic commitment of an autonomous will is to act in accordance with whatever intermediate laws can be willed by any and all autonomous wills as standards that must be followed by everyone whether they want to or not. Therefore, an autonomous agent is disposed to act only in ways that are justifiable to other autonomous agents.

Self-preservation, exercise and reciprocity are therefore constitutive features of autonomy, so if rational agents necessarily have autonomy of the will then they are constitutive features of rationality as well. If this argument is correct, rational agents are by their nature autonomous and so disposed to preserve themselves, exercise their rational capacities and reciprocate with others. These features of rational agency are substantive rather than formal, but they are also quite thin because self-preservation, exercise and reciprocity are about abstract rational, autonomous nature itself independent of its embodiment in finite human beings.

Substantive rationality in commonsense

The argument from modest constitutivism may not persuade those who are skeptical of Kant’s larger project of vindicating reason and its normative authority. If we set aside his more ambitious aims, however, and focus on the normative account Kant and his followers give of a reason-governed agent, we find that broadly Kantian conceptions of the person fit fairly well with commonsense. The second way of arguing for a substantive account of rationality is simply to lay it out as a normative starting-point, show how it connects with everyday normative thinking, and then build a normative theory from it that we would endorse as a whole on due reflection.

This is basically the same holistic strategy that I argued substantive realists about reasons, such as Scanlon, employ to justify taking a wide variety of unexplained reasons as normatively primitive. When Scanlon claims that he is realistic about reasons, he is not endorsing a metaphysical view about the nature of reasons (whether realist or anti-realist); instead, he believes that within the normative domain, the buck stops with reasons, they are foundational in the sense that there is no further normative explanation for why facts are reasons.  

32 Scanlon supports his starting points, not with meta-normative arguments or straightforward normative
ones, but by claiming that SRR both fits with commonsense judgments about reasons and also helps to explain a variety of other normative notions, such as desires, values and moral principles, as part of a comprehensive normative theory that he thinks can secure our allegiance.

Defenders of KCR should follow Scanlon’s lead by starting with a different reasonable starting point grounded in common normative understanding, without getting much into its meta-ethical credentials, and try to work out a more unified normative theory on that basis. But instead of beginning with a hodge-podge of reasons, advocates of KCR should clarify their conception of a rational agent and show how it can capture the reasons we recognize in commonsense. The success of the overall normative structures in which SRR and KCR figure provide some of the most compelling support for SRR and KCR themselves.

One advantage of adopting this kind of normative holism is that constructing a normative theory on the basis of SRR or KCR becomes somewhat less daunting. Capturing all the reasons of commonsense will seem particularly implausible if a defender of KCR can only appeal to her abstract and often vague intuitions about the nature of rationality itself. Allowing her judgments about particular reasons, which are often more vivid and deeply held, to influence her understanding of rationality may seem circular when she turns around and claims that, from the standpoint of her normative theory, those reasons are explained in terms of rationality. However, if her goal is general and wide reflective equilibrium, then considered normative judgments at all levels of generality, from ones about rationality to those about specific reasons, are relevant pieces of evidence that she is trying to capture as she builds her normative ethical theory. Suppose she postulates, as a provisional but central component of her theory, that anytime there is a genuine reason, there must be an explanation of it in terms of rationality. As she tries to fill in the details of her view, she will consult her intuitions about the nature of a rational and reasonable person, but there will inevitably be reasons she believes exist that so far do not seem to have a basis in rationality. When there is an apparent conflict between her theory of rationality and her judgments about particular reasons, holism gives her a number of options: She could abandon her judgment that this fact is a reason, she could change her conception of rationality to accommodate the reason, or she could give up her “load bearing” supposition that all reasons are explained in terms of rationality. Which option she should take depends on which of the resulting normative theories as a whole is more likely to stand in reflective equilibrium. But it would not be circular for her to alter her conception of rationality to accommodate the reason because, as she sees it, her normative theory explains the reason in terms of rationality while her considered judgment that the reason exists merely alerted her to the presence of that explanation. Similarly, when a mathematician finds a specific claim plausible, this may set her looking for a proof of it that, if found, explains why the claim is true. And, Rawls credits his account of the original position to intuitions about fair procedures and also to judgments about the justice and injustice of specific social policies, even though he thinks that those policies are justified or not by the original position itself. Defenders of KCR can therefore appeal to intuitions about particular reasons in constructing their theories, while supporters of SRR can, eo

---

33 Rawls (1999b, pp. 33-35) takes the wrongness of slavery as a provisional fixed point and spends Part II of A Theory of Justice drawing out some implications of his account of justice as fairness that, if they fail to secure our allegiance, may require fiddling with some of the conditions of the original position.
ipso, do the same for intuitions about rationality, which can alert them to the presence of certain reasons.  

Suppose our aim is reflective equilibrium and suppose we take KCR as a core working assumption. As we attempt to specify a suitable conception of rationality for our normative theory, four kinds of considered judgments will be relevant. All three types support, or at least do not undermine, incorporating self-preservation, exercise and reciprocity as substantive elements of rationality.

(1) We have considered judgments from ordinary language about the meaning of ‘rationality’. The concept that word picks out, apart from particular conceptions of rationality, has several features that distinguish it from other concepts, such as impulse, instinct, feeling, perception and faith. These analytic constraints restrict what counts as a conception of rationality, as opposed to conception of something else. Following Rawls, we can say that issues of meaning have a limited but real role to play in normative theory, although they do not determine which of the candidate ways of interpreting a normative notion is correct – that issue is settled by which of the eligible interpretations matches our considered judgments in reflective equilibrium.

Here are five analytic features of rationality, none of which exclude conceptions of rationality that incorporate non-formal features. First, requirements of rationality are valid for all agents. Second, agents are criticizable if they violate requirements of rationality. Third, the most basic requirements of rationality are necessary and free of contingencies. Fourth, we expect stable convergence in the judgments and actions of fully rational agents when they reason about a wide variety of matters on the basis of the same facts, and such reasoned conclusions are worthy of our acceptance. Fifth, fully rational people reason on the basis of mutually recognized rules and evidence and reach agreement on the basis of deliberation, discussion and judgment rather than propaganda, rhetoric or prejudice.

None of these essential elements excludes the possibility that a conception of rationality can include self-preservation, exercise and reciprocity – those putative requirements are meant to be universally and necessarily binding, they are supposed to constitute grounds for rational criticism, and they can support stable convergence among agents who reason together on the basis of these mutually recognized principles.

Someone who is opposed to incorporating substantive elements of any kind into a conception of rationality could try to argue for a further conceptual feature of rationality, namely that rationality is formal as well, concerned only with consistency and coherence among mental states. On this view, there is no rational assessment of final ends regardless of how wrong, vicious, wacky or unjustified they may be. Rationality requires us to avoid contradictory beliefs and take the believed necessary means to our ends or give them up, for example; a fully rational person organizes and schedules her ends efficiently; and rationality may require us to have whatever desires we believe we have most reason to have or to abandon those judgments. Formal rationality cannot generate reasons to have particular mental states because its requirements are conditional, they leave us with options about how to resolve incompatibilities.

---

34 Jonathon Dancy (2014) seems to take this approach when he argues that intuitions about the instrumental principle of rationality alert us to the presence of particular reasons to be prudent, which explain why that principle itself holds.
35 Rawls (1999f, pp. 114-119) claims that there are formal constraints on what counts as a conception of right, including generality and universality.
36 (Broome, 2010; Rawls, 1999f; Scanlon, 1998)
among them – all things equal, we are no less rational for abandoning one contradictory intention over the other, for instance. What we have substantive reasons to do is therefore treated as a separate matter from good mental house-keeping.

Rationality may be purely formal, but we should not settle this matter on conceptual grounds alone but instead treat formal ways of thinking about rationality as candidate conceptions of rationality and assess them holistically against those that incorporate non-formal elements. For one thing, certain principles of inductive and abductive reasoning are plausibly regarded as features of rationality even though some of them are not altogether formal. Second, if we are clear to distinguish the notion of rationality as it figures in, for example, rational choice theory from a more general idea of rationality as ‘having to do with reason’ then we can admit that, as a conceptual matter, the former idea may be formal while the latter idea may involve substantive elements such as self-preservation, exercise and reciprocity.

(2) A second class of considered judgments we have about rationality comes from philosophical tradition. A prominent issue in the history of moral philosophy has been about how to understand reason, how it differs from sentiment, what role it plays in deliberation, how it relates to morality, and so on. Plato, Aristotle, Aquinas, Hobbes, Hume, Leibniz and Kant had very different views about rationality, some of which were exclusively formal while others were substantive and thick, but arguably they were attempting to capture and explain the same rational aspect of human nature. On an oversimplified picture, Plato, Aquinas and Leibniz thought that rationality includes a capacity for rational insight into an external moral order. That view strikes many of us as implausible, but not primarily because it offends our conceptual intuitions; rather, we recognize good metaphysical, epistemological and normative considerations against it. But if we set aside rational insight, we also find rationalists claiming that rationality involves substantive rational dispositions to act. Reason, on this view, can generate its own motivational states and, like a charioteer, guide, control and oppose our non-rational desires. Aristotle extended the idea of rationality even further, claiming that a person who is fully governed by reason, in virtue of her nature as such, possesses the intellectual virtues, exercises practical wisdom well and has a strong will. And Kant thought that we are rationally required to preserve ourselves as rational agents, develop and exercise our rational capacities and, above all, reciprocate with others by acting only on maxims that we can will as universal law. Hobbes and Hume, of course, opposed the rationalist tradition and advocated formal conceptions of rationality in place of substantive ones.

These philosophers had subtle and complicated views about the nature of reason that should guide us as we attempt to formulate a conception of rationality that matches our considered judgments on due reflection. Substantive conceptions of rationality should not be dismissed out of hand as already refuted or as novel but untested, for there is a long history within normative theory to thinking about rationality as more than just consistency and coherence among mental states. On the other hand, substantive realism about reasons is a relative newcomer to normative theory.37

(3) A third class of considered judgments we should consult are commonsense ones involving a family of related ideas, including reason, reasons, reasoning, reasonableness,

---

37 Broome (2004, p. 28) describes the 1970’s as the “age of the discovery of reasons” and there were so-called ‘good reasons’ theorists, such as Baier (1958); Hampshire (1949); Toulmin (1950), working shortly before then.
rationality and rationale.\textsuperscript{38} These everyday intuitions are relevant to understanding rationality in the broadest sense of ‘based on reason’.\textsuperscript{39}

A cursory glance at the ways we typically think and talk about these cognates in commonsense suggests that rationality is not just limited to consistency and coherence among mental states but also includes prudence, sensitivity to evidence, open-mindedness, strength of will and fairness. We talk about a mob recovering their reason before going on a rampage, we try to reason with a clearheaded person who is about to commit a crime, we say that certain psychopaths are beyond reason, and we reject someone’s rationale as biased and prejudicial. Thomas Jefferson wrote: “A government of reason is better than one of force” and Abraham Lincoln later agreed: “Reason, cold, calculating, unimpassioned reason, must furnish all the materials for our future support and defense. – Let those materials be moulded into general intelligence, sound morality, and in particular, a reverence for the constitution and laws.”\textsuperscript{40}

Some of these ideas of reason have been codified in the law: We have the “reasonable person” test, which considers how a conscientious, appropriately informed, fair-minded person would act in cases in which she may harm others, the Supreme Court has been described as having an “aura of reason”\textsuperscript{41}, and we require juries in criminal cases to find someone guilty “beyond a reasonable doubt.” We say that a colleague’s ranking of job candidates has the weight of reason behind it and that a price is reasonable; we implore others to be reasonable, we think it is irrational to give no weight to one’s future wellbeing, and we say that the administration’s reasoning is flawed because it did not take proper account of the interests of some people. And we think that there are good reasons to be generous to our neighbors and that it stands to reason that they should reciprocate.

When we draw on this complicated family of judgments in commonsense, we find good reasons for thinking that self-preservation, exercise and reciprocity are requirements of rationality:

Consider, first, our everyday notion of a reason-governed person. She not only has consistent beliefs and purses her ends efficiently, but such a person is also sane, judicious and sensible, exercises good judgment, and treats others fairly. We call someone irrational in this more robust sense if she is senseless, stupid, unfair, unreasonable or lacking sound judgment. If someone really does not care at all about her life and makes no effort to protect it, we would tend to say she is irrational for not caring about her own good. If she recklessly endangered her life for the mere thrill of it, attempted suicide out of passing grief, or allowed herself to be killed for another’s enjoyment, we would likely regard her as unreasonable for not having proper respect for oneself. There is something rationally criticizable, in a broad sense, about a person who is indifferent to her continued existence, who would just as soon perish as live. We would tend to think such a person lacks good sense, and we might express our concern by imploring an extreme sports enthusiast or a person contemplating suicide out of self-contempt to be reasonable, although we may not be so quick to judge as unreasonable someone who sacrifices

\textsuperscript{38} One can certainly argue that apparent similarities mask deep underlying differences – in spite of their common etymologies, for example, John Broome claims that “the connection between rationality and reasons is not very close” (Broome, 2010: 288) because being a rational person is a matter of good mental house-keeping while reasons tend to be substantive considerations that count in favor of adopting attitudes.

\textsuperscript{39} So-called ‘good reasons’ theorists of the 1940’s and 1950’s were arguably attempting something similar. See (Baier, 1958; Hampshire, 1949; Toulmin, 1950)

\textsuperscript{40} (Jefferson, Lipscomb, Bergh, Johnston, & States, 1905, p. 15:284; Lincoln, Basler, & Abraham Lincoln Association (Springfield Ill.), 1953, p. 1:50)

\textsuperscript{41} (Dowd, 2012)
herself to save others or to end her own excruciating pain. Similarly, if someone did not care at all about exercising her rational capacities, but was content to allow her powers of mind to waste away, we would not regard her as fully reasonable. And, our commonsense idea of a reasonable and rational person is not that of a complete altruist, but nor is she wholly self-interested; instead, a person who is fully governed by reason is fair, she makes reasonable demands but does not take excessive advantage of her bargaining position, she is willing to do her fair share and abide by shared standards as long as others are willing to do the same even when doing so requires her to sacrifice some of her interests. Our idea of a person of reason therefore also includes a basic idea of reciprocity, mutuality and giving the same to others. These considerations suggest that self-preservation, exercise and reciprocity are requirements of reason.

Second, we clearly recognize in commonsense particular reasons for agents to care about their continued existence as such, to exercise their rational powers, and to avoid acting in ways that are not justifiable to others. If KCR is correct, these considerations are reasons in virtue of the attitudes, commitments and choices of rational agents. Therefore, working backwards from particular reasons to rational nature, we find that rational agents, in virtue of their nature as such, are rationally committed to preserving themselves, to exercising their rational powers, and to acting only in ways that are justifiable to one another. This means that self-preservation, exercise and reciprocity are requirements of rationality, on the working assumption that reasons are explained in terms of rationality.

A broadly Kantian contractualist moral theory

Even if we allow a conception of rational agency that includes a few substantive yet thin features such as self-preservation, exercise and reciprocity, it is still not clear whether that idea is robust enough to capture the reasons we recognize in commonsense. In response to the second horn of the dilemma for KCR, I have argued that substantive conceptions of rationality can count as genuine, but the first horn requires giving some sense of how a substantive conception of rationality along these lines can generate all of the reasons that seem evidently correct – the proof is in the pudding, as it were.

It is very important to separate out the kinds of reasons we are trying to capture with the idea of rationality, such as those of personal virtue, interpersonal morality, justice, prudence, the environment, animals and aesthetics. We should consider each domain individually because the role of rational agency may not be exactly the same in explaining each type of reason. Reasons about the arrangement of basic social institutions, for example, may be explained by the choices of rational agents who take up the point of view of Rawls’ original position; reasons of prudence may best be explained by principles of instrumental rationality; and reasons provided by non-moral values may be explained in terms of the attitudes and choices of fully rational and reasonable human beings.

To illustrate how one such explanation might go, consider what kinds of moral reasons can be generated by a specific kind of contractualist procedure in which rational agents who are committed to self-preservation, exercise and reciprocity settle on mid-level moral principles that are acceptable to all.

As a type of constructivist theory, KCR is not limited to a subsumptive model of justification in which reasons follow logically from principles of rationality along with non-moral facts, although some reasons may be generated in that way – for example, rational agents are disposed to preserve themselves, but in our world having sufficient nutrition is necessary for a rational human being to persist, so such people have reasons to acquire enough food to survive. KCR can rely on procedures of construction as well, in which rational agents take up a specified
point of view from which they work out, on their own or together, solutions to practical problems on the basis of deliberation and judgment. We then regard their deliverances as true or reasonable because they result from a justified procedure of construction, and we regard as reasons those facts that weigh in their deliberations.

The metaphor of construction raises questions about what is to be constructed, what materials are to be used, and who is doing the constructing – the problem is to specify, in a non-arbitrary way, the motivations of the parties, the information they have available and the context in which they make their decisions. As Rawls’ list of alternative descriptions of the original position illustrates, there are various options available for each element, and the suitability of any one depends on its inherent plausibility, the way it helps to generate plausible results when the procedure is applied, and the success of the overall normative theory in which it figures.

Here is one kind of broadly Kantian contractualist moral theory that is different from Scanlon’s, in which rational agents settle on mid-level moral principles that are acceptable to all on the basis of their self-regarding rational commitments.

At the first stage of the theory, the subject of the agreement is mid-level principles of interpersonal morality that apply to all rational agents. The parties are motivated by the self-regarding rational commitments that are constitutive of their nature as rational agents. The parties are not moved by subjective characteristics, desires or inclinations, nor are they moved by reasons that are independent of their rational nature. And their decisions are made on the basis of information that is relevant to the task of specifying universal rational standards.

Suppose, for example, that rational agents are presumptively committed to preserving themselves as rational agents. Each rational agent in the procedure would presumptively insist that his life be preserved, so any candidate principle that allows certain kinds of killing would be rejected by those who stand to lose because of the rational interest they necessarily take in their own survival. Therefore, the parties would settle on a presumptive principle against killing one another. Similarly, contractors are presumptively committed to exercising their rational capacities, so principals that denied this opportunity to some or did not ensure the conditions necessary for everyone to do so, would likely be rejected by some, which means they would choose a presumptive principle against thwarting or undermining anyone’s ability to exercise her rational capacities. We do not start out building in a prohibition on certain kinds of killing, deception or manipulation in our conception of rational agency; instead, we begin with substantive but thin self-regarding interests that rational agents have in virtue of their nature as such and combine them with their overriding commitment to reciprocate with one another to construct other-regarding moral requirements.

The first stage of the theory is likely to generate mostly abstract presumptions that offer limited practical guidance. Although this result seems appropriate, because the principles chosen at this level are supposed to apply to all rational agents, we can define further stages of the contractualist procedure to interpret and apply the most basic requirements. In a moral contractualist theory, levels are defined by the sorts of questions that the parties are to consider, the type of information about themselves and their world that they are permitted to take into account and their motivations in addressing such issues. Lower levels in the theory are constrained by the results of higher ones.

At the second stage, the subject of the agreement is principles of interpersonal morality that apply to all limited, embodied and dependent rational creatures living in a world like ours. Their motivations are self-regarding rational dispositions that are constitutive of their nature as

---

42 (Rawls, 1999f, pp. 126-127)
rational as such, along with any necessary and useful means in our world to satisfying these commitments. Here it becomes relevant that the parties have bodies that make them vulnerable in various ways, that food, shelter and medicine are scarce, and that we may sometime need others to save our own life. And the parties have access to any information that is relevant to this more specific task. At each level, the parties are making decisions in virtue of their nature as rational gents, but contingent, accidental and specific facts about our world will certainly figure in their deliberations at lower levels as they give further content to more specific moral principles and try to resolve conflicts among them. Not only would the parties at this level settle on more specific principles about mutual aid, endangering the food supply, etc., they may also begin to make exceptions to certain presumptive rational prohibitions. The contractors might choose, for example, to allow assisted suicide in cases in which a person’s ability to function as a rational agent is tenuous and likely to worsen. Some of the parties may regard their interest in exercising their rational powers as more important than their interest in preserving themselves, so they would endorse a principle allowing suicide in order to keep their rational capacities from degenerating much further. As in the first stage, they would settle on these principles in light of their self-regarding rational interests, but here these interests are given further content by human nature and the natural world they occupy.

There may even be further stages in which contractors consider more specific issues that arise in particular times, societies or groups, such as what some community should do about foot shortages, how much life-saving help members of our society are required to give one another, how much risk fast-driving ambulances can impose on drivers, etc. At each level, the parties are making decisions in virtue of their nature as rational agents, but contingent, accidental and specific facts about our world will certainly figure in their deliberations at lower levels as they give further content to more specific moral principles and try to resolve conflicts among them.

I have developed aspects of this contractualist moral theory in more detail elsewhere, but these examples give some idea of how KCR can move from the very abstract and unified to the very specific and particular, and back again. A defender of SRR, who remains skeptical of expanding the notion of rationality, could even adopt a version of this contractualist theory as a way of structuring the particular reasons that individuals have, whether or not those reasons can be unified in the notion of rationality. But we can also see more clearly what proponents of KCR find most troubling about Scanlon’s moral contractualism and why they are so intent on specifying a conception of rational agency that can generate the moral reasons that exist. As Kantians see it, morality is fundamentally self-legislated rather than imposed on us by God, self-interest, sentiments or moral considerations that exist independent of our autonomous wills – by affording such a prominent role to SRR, Scanlon’s moral theory is heteronomous. If, however, there are self-regarding interests that rational agents have as part of their nature as such then we can explain what moves them to accept or reject mid-level principles in terms of the nature of autonomy itself.

In sum, KCR can distinguish between the standpoint of ordinary people on the streets who make particular normative judgments, from the perspective of reflective people who are thinking about the more general principles, values and facts that provide reasons for or against particular judgments. Then there is the perspective of the normative theorist who proposes a constructivist procedure to try to account for and justify mid-level normative principles and ultimately the specific reasons of everyday life. Along the way she attempts to justify her

43 [Citation suppressed for blind review]
procedure itself, both by the plausibility of its constitutive features and its success at generating the kinds of conclusions that we would endorse on due reflection.

Works cited


