Journal for the History of Analytical Philosophy

Volume 4, Number 8

Rawls on Kantian Constructivism

Nathaniel Jezzi

John Rawls's 1980 Dewey Lectures are widely acknowledged to represent the *locus classicus* for contemporary discussions of moral constructivism. Nevertheless, few published works have engaged with the significant interpretive challenges one finds in these lectures, and those that have fail to offer a satisfactory reading of the view that Rawls presents there or the place the lectures occupy in the development of Rawls's thinking. Indeed, there is a surprising lack of consensus about how best to interpret the constructivism of these lectures. In this paper, I argue that the constructivism presented in the Dewey Lectures is best understood as involving the view that moral truth is correspondence with procedurally-determined, stance-dependent facts. Employing Rawls's discussion of rational intuitionism as a foil, I defend this reading against textual discrepancies from within the lectures, as well as those one finds across Rawls's other works. In addition to settling interpretive disputes, I draw out the ways in which this understanding of Kantian constructivism fits within the broader comparative project in 'moral theory' that Rawls inherits from Sidgwick.

Rawls on Kantian Constructivism

Nathaniel Jezzi

1. Introduction

Many philosophers have shown an interest in constructivism in ethics in recent years—either presenting (Hill 1989; Darwall, Gibbard and Railton 1992; O'Neill 2003a,b), or defending (Rawls 1980; Scanlon 1982; O'Neill 1989; Milo 1995; Korsgaard 1996, 2003; James 2007, 2012; Street 2008, 2010; LeBar 2008; Lenman 2010, 2012; Hill 2011; Dorsey 2012; Bagnoli 2013a; Silk 2015; Schafer 2015a,b), or critiquing (Brink 1987, 1989; Timmons 2003; Gibbard 1999; Shafer-Landau 2003; Fitzpatrick 2005; Hussain and Shah 2006; Enoch 2009; Hussain 2012; Ridge 2012) versions of the view. The list of participants in this debate continues to grow, and there is still currently no consensus on what constructivism is supposed to be or how exactly it relates to realism or other available metaethical positions, like response-dependence theories or expressivism (Street 2010, 364). However, there is a broad consensus amongst these philosophers that constructivism in ethics first finds contemporary expression in John Rawls's 1980 Dewey Lectures, collectively titled 'Kantian Constructivism in Moral Theory' (hereafter *KCMT*).[1] In fact, nearly every major recent contribution to this debate cites *KCMT* as the *locus classicus* for contemporary discussions.

Nevertheless, few philosophers have engaged explicitly with the details of the view Rawls presents there, perhaps because there are significant interpretive challenges that must first be overcome.[2] In particular, Rawls makes varied remarks about the role of moral truth in constructivism and the independence of moral theory from other areas of philosophical investigation. While these remarks fit well with his characterization of justice as fairness in later works, it is less clear how one is to square them with his characterization of constructivism in the lectures. Indeed, some might reasonably question whether there is a productive and coherent way of accommodating everything that Rawls says in the lectures or whether a careful reconstruction is worth the effort, especially when this is arguably not essential to understanding the trajectory of his larger and more central project in political philosophy.

One of my aims in this paper is to defend a careful, textually-based reading of the constructivism of *KCMT* against such challenges, one that should afford us with a clearer historical backdrop to current presentations and debates. The interpretation I defend is one according to which Rawls's Kantian constructivism in the lectures is to be understood as involving (if not committing one to) a form of stance-dependent cognitivism.[3] On the one hand, as a view in moral theory, Kantian constructivism asks us to entertain what it would be like to regard ourselves *as if* such a stance-dependent cognitivism were true—from within an engaged, first-order perspective, so to speak—and to work out how this form of self-regard would structure norms for relating to ourselves and others. On the other hand, Rawls allows

---

[1] As Brian Barry (1989, 264–65) recognizes, it is also the first place that the term 'constructivism' shows up in Rawls's work.

[2] Brink's (1989) discussion is the notable exception. The same account can also be found in the earlier Brink (1987). However, as Brink's later discussion is 'essentially unchanged' and also more comprehensive than the earlier discussion (Brink 1989, 303n), I will cite the later text in further reference to Brink's views. This is not to say that other philosophers have not published claims about the role of *KCMT* in the development of Rawls's thinking about justice. See, for example, Barry (1989), Freeman (1999, 2007) and Weithman (2010). However, for all of their insight, none of these works offer a sustained, close reading of the lectures in the way that Brink does, or that I attempt here.

[3] This distinction between involving and committing will be explained in sections 3 and 4.

that the conclusions we draw within this engaged, first-order perspective bear on our theoretical commitments outside of this perspective.[4] To be clear, my claim is not that Rawls intended to present the view this way or that this characterization of constructivism represents his own considered view at the time of writing these lectures;[5] rather, what I offer is a rational reconstruction of the view one finds there on the basis of philosophical and textual coherence.[6]

Another related aim is to draw attention to the way in which the view that Rawls develops in the lectures contributes to the broader comparative project in 'moral theory' that he inherits from Henry Sidgwick. Although Rawls holds up Sidgwick's work in *The Methods of Ethics* (hereafter *Methods*) as a model for his own approach to moral theorizing, he also thinks that

Sidgwick fails to fully appreciate the structural complexity of the task. By identifying exactly why Rawls thinks that Sidgwick's framework for moral theory fails and how he tries to correct for this, we will better appreciate the relevant comparison class for Kantian constructivism, along with the details of this view and its further commitments. What's more, this discussion will reveal that the standard way of situating *KCMT* within Rawls's work obscures the lectures' real role and significance within the development of Rawls's thinking. Instead of reading the lectures as a 'transitional stage' between better known works, I argue that they represent the culmination of Rawls's response to Sidgwick and that, in this respect, we better appreciate their importance in relation to an altogether different set of texts.

Although I believe that the reading I offer is the best way to make sense of all of the competing evidence one finds across the text, the point of this exercise is not merely to settle interpretive disputes. The view I uncover also reveals something important and, as of yet, underappreciated about both the structure of and possibilities for an approach to moral theorizing that assigns primacy to conceptions of moral personality.

The structure of this paper is the following: In section 2, I present a 'conventional' reading of *KCMT* alongside the two interpretative challenges it faces: Rawls's disavowals of moral truth and metaphysics. In section 3, I contrast Rawls's conception of moral theory with the approach Rawls finds in Sidgwick and explain what this reveals about the relevant comparison class for Kantian constructivism. In section 4, I present the details of Kantian constructivism within this framework, drawing special attention to the way Rawls understands the Kantian conception of autonomy, and explain how the resulting view requires us to reframe the role of the conventional interpretation in Rawls's thinking. In section 5, I return to the interpretive challenges I introduce earlier and argue that a response is not to be found in familiar works but, rather, in a heretofore unappre-

---

[4]Here, an instructive contrast can be found in the quasi-realist approach of some contemporary expressionists. See, e.g., Blackburn (1998) and Gibbard (2003). Like quasi-realism, Rawlsian moral theory invites us to consider metaethical ways of thinking and talking from within the standpoint of an engaged, first-order perspective. However, unlike quasi-realism, the aim is not to neutralize or quarantine this way of thinking and talking within the first-order perspective; rather, Rawls wants to allow that our first-order reflection on metaethics might have consequences for our second-order, metaethical commitments. Following Mark Timmons, we might think of Rawlsian moral theory as part of a project of *internal accommodation* whereby 'a plausible metaethical view should comport with deeply embedded assumptions of ordinary moral discourse and practice' (1999, 12). Both Korsgaard (2003) and Street (2010) have acknowledged ways in which their own versions of constructivism share features with expressivism. My reading of *KCMT* suggests a way in which Rawls's own constructivist project in moral theory might also share something with the approach taken by defenders of these views.

[5]Rawls is, in fact, quite explicit that he is *not* defending Kantian constructivism in the lectures but, merely, explicating this position for the purposes of moral theory (*KCMT*, 570).

[6]Brink is also explicit that his reading is to be understood as a *reconstruction* in this sense. See, e.g., Brink (1987, 72–73), and compare Weithman (2010, 21). However, I think that when we apply Brink's own interpretive methods to the text we see that his reading cannot be correct. Rather, the correct view involves the stance-dependent cognitivist reading I defend here.

ciated source, viz., Rawls's doctoral dissertation, and that what Rawls says there supports my interpretation of *KCMT*. In section 6, I conclude by highlighting how the arguments I have presented require us to resituate *KCMT* within the broader context of Rawls's writings and reconceive its role in the development of his thinking.

2. The Conventional Interpretation and Its Challenges

Constructivism is a broad family of views. However, as many defenders of constructivism find inspiration for their views in Kant's moral philosophy, it is understandable that, up until quite recently, discussions about constructivism have centered on Kantian varieties.[7] Although there is presently no uncontroversial understanding of *Kantian constructivism* in ethics,[8] certain interpretations have enjoyed more support than others. In particular, the view is often presented, or interpreted, as a metaethical view that combines the following three claims:

I. Moral statements are truth-apt.
II. At least some moral statements are true.
III. What makes these statements true are facts about what rational agents would choose, or agree to, in a choice procedure that represents persons as free and equal.

Let us call this the *Conventional Interpretation*. Claim I distinguishes constructivism from expressivism and traditional forms of non-cognitivism. Claim II marks the view as a version of a success theory (i.e., a rejection of Mackean error theory). Claim III states that the truth-making facts for moral statements are, in

the words of Ronald Milo, *stance-dependent*—i.e., 'they consist in the instantiation of some property that exists only if some thing or state of affairs is made the object of an intentional psychological state (a stance), such as a belief or a conative or affective attitude.'[9] In the case of Kantian constructivism, moral facts depend on our *volitional* stances, or what we would choose, or will, or agree to in the course of the procedure. This last condition serves to distinguish the view from certain forms of *stance-independent* moral realism (i.e., views according to which the nature of moral facts is, at bottom, independent of what we desire, believe, choose, will, or agree to)—such as rational intuitionism—on the one hand,[10] as well as certain stance-dependent cognitivist views—such as ideal observer theories and extreme versions of metaethical relativism—on the other hand.[11]

Although Rawls discusses Kantian constructivism in several of his works,[12] one finds his most extensive and detailed presen-

---

[9]See Milo (1995, 192). Both Shafer-Landau (2003) and Street (2008) follow Milo in describing the constructivist as committed to a stance-dependent view of moral facts.

[10]For coinage of this term, see Milo (1995, 182).

[11]One may object that III, in referencing 'rational agents' or 'persons as free and equal', smuggles moral/normative content into the account, rendering the view first-order (cf. Hussain and Shah 2006; Hussain 2012). For some recent support for interpreting Kantian constructivism as a view in metaethics, see Ridge (2012, 140); Galvin (2010, 22ff). Note, however, that even if Kantian constructivism counts as a view in metaethics, there is the further question of whether it counts as a novel, interesting, and free-standing alternative to existing metaethical views. For some doubts about this see Hussain and Shah (2006); Hussain (2012); Street (2008); Enoch (2009); Ridge (2012). Although I mostly set this further question aside, there is a way in which the reading I offer here helps to mitigate this concern. For if, as I argue, Rawls's aim in introducing metaethical considerations is to characterize competing self-conceptions and related forms of self-regard from within moral theory, the further question as to whether these metaethical considerations constitute a determinate and free-standing position in metaethics is largely beside the point.

[12]See Rawls (1985, 230–31, 237n, 239n, 246–47); (1989, 81–113, 95–102); (1993/1996, 99–101, 110–18); (2000, 235–52). As I will argue in due course,

---

[7]More recently, there has been a proliferation of constructivisms being defended and debated in the philosophical literature—including Aristotelian (e.g., LeBar 2008), Humean (e.g., Lenman 2010; Street 2008, 2010) and Nietzschean versions (e.g., Silk 2015).

[8]See Galvin (2010) for a survey of recent positions.

tation of this view in *KCMT*. As in earlier work, Rawls's concern in these lectures is to explicate and ground his account of justice as fairness—'the idea that the principles of justice are agreed to in an initial situation that is fair' (Rawls 1971/1999, 11; compare *KCMT*, 552). However, Rawls is clear that justice as fairness only represents one particular 'variant' (*KCMT*, 515) or 'illustration' (*KCMT*, 535) of Kantian constructivism in moral theory. The following oft-quoted passages are instances of Rawls's characterizing Kantian constructivism more broadly.

> Kantian constructivism holds that objectivity is to be understood in terms of a suitably constructed social point of view that all can accept. Apart from the procedure of constructing the principles of justice, there are no moral facts. Whether certain facts are to be recognized as reasons of right and justice, or how much they are to count, can be ascertained only from within the constructive procedure, that is, from the undertakings of rational agents of construction when suitably represented as free and equal moral persons. (*KCMT*, 519)

> The parties in the original position do not agree on what the moral facts are, as if there already were such facts. It is not that, being situated impartially, they have a clear and undistorted view of a prior and independent moral order. Rather (for constructivism), there is no such order, and therefore no such facts apart from the procedure of construction as a whole; the facts are identified by the principles that result. (*KCMT*, 568)

These passages provide evidence for thinking that Rawls understands Kantian constructivism to be the view that *moral* principles (or 'reasons of right and justice') are determined by the hypothetical agreements that follow from his favored construction procedure—viz., *the original position*, 'a situation that is fair

there are important differences between Rawls's discussion of 'the Kantian interpretation of justice as fairness' in §40 of Rawls (1971/1999, 221–27) and the constructivism of *KCMT* and later writings; the two views should not be conflated. This point has not been sufficiently appreciated by commentators (including, e.g., Freeman 1999, 2007; Weithman 2010) and has led them to misconstrue the role of *KCMT* in the development of Rawls's thinking.

between [the parties] and in which they are represented solely as free and equal moral persons' (*KCMT*, 522; see also *KCMT*, 519, 524, 543, 548, 559, 560–61, 564, 565.) These statements appear to confirm an understanding of constructivism that would include all three of the conditions presented above: there are moral *facts* which serve as truth-makers for moral statements, but these facts are not stance-independent; rather they are picked out by procedurally-dependent principles—i.e., principles that the parties to the procedure would *agree to*. So, the Conventional Interpretation would appear to capture the constructivism of *KCMT*.[13]

Although there is good *prima facie* reason for reading the constructivism of *KCMT* this way, one does not have to look any further than *KCMT* itself to find evidence that complicates the claim that Rawls's constructivism is a stance-dependent, cognitivist success theory. There are two main complications that must be overcome by anyone who would apply the Conventional Interpretation to the constructivism of *KCMT*.

First, throughout the lectures, Rawls repeatedly insists that Kantian constructivism makes no claims to moral truth. To-

[13]One will notice that the characterization of constructivism I present here presupposes (or at least suggests) a correspondence theory of truth and a correlative ontology of truth-makers. However, it is not clear that the Conventional Interpretation requires this. Rather, it would appear possible for a defender of this interpretation to remain agnostic on the nature of truth or even to accept talk of 'truth-makers' and 'facts' but deny that this commits one to any form of metaphysical realism or substantive explanation. Moreover, one might argue, the available textual evidence would also not appear to force a correspondence theory upon Rawls. I concede both of these points. As will become clear later in the discussion, my point is only to show just how much metaphysical weight Kantian constructivism in moral theory can bear. Although nothing in the text requires us to read constructivism as involving a correspondence theory, there is also nothing that Rawls says that precludes this. What's more, there is even some reason to think that a correspondence theory would serve to strengthen the contrast that Rawls intends to draw between rational intuitionism and constructivism. Thanks to an anonymous referee for helping me to clarify this point.

wards the beginning of the first lecture, Rawls notes that in constructivism 'the search for reasonable grounds for reaching agreement rooted in our conception of ourselves and in our relation to society replaces the search for moral truth' (*KCMT*, 519). He repeats this point again in his concluding remarks in 'Lecture III'.

> This rendering implies that, rather than think of the principles of justice as *true*, it is better to say that they are the principles most *reasonable for us*, given our conception of persons as free and equal, and fully cooperating members of a democratic society. (*KCMT*, 554, emphasis added)

This shift away from talk of the truth of moral principles to talk of their reasonableness may appear confusing. Why does Rawls introduce *reasonable for us* as an alternative to truth? Does this shift suggest that we ought not to interpret Kantian constructivism as a version of moral cognitivism? This disavowal of truth presents an apparent challenge for the three-point Conventional Interpretation, since each point of its characterization of Kantian constructivism invokes truth.

Second, and perhaps relatedly, Rawls appears to describe *justice as fairness*—the variant of Kantian constructivism he is most concerned with in *KCMT*—as independent of metaphysics and other metaethical concerns. In fact, in later works, he clearly distances justice as fairness from the kind of metaethical constructivism I have described above, opting instead for a metaphysically neutral 'free-standing conception' of justice (see, esp., Rawls 1993/1996). Whereas the Conventional Interpretation involves metaethical commitments, these remarks suggest that Rawls's view is not to be interpreted as a metaethical doctrine at all.

Despite Rawls's apparent disavowal of truth in *KCMT* and clear distancing of justice as fairness from metaethical constructivism in later works, I will argue that the view Rawls presents in the earlier *KCMT* ought to be understood as involving an appeal

(if not a commitment) to the above three claims and, indeed, that this represents the most plausible and coherent reading of the view. If I am correct, the Conventional Interpretation ends up getting something right. But it also counts as mistaken, or misleading, in the sense that it fails to adequately capture the complexity of Rawls's view. In order to appreciate the interpretive arguments I will present, it will first prove useful to have an understanding of what Rawls means when he describes Kantian constructivism as a position *in* moral theory. I will address this point over the course of the following two sections.

3. Advancing Sidgwickean Moral Theory

While it would be hard to miss the influence of Kant on Rawls's discussion in *KCMT*, few have commented on the extent of Sidgwick's influence.[14] This is quite surprising, considering that Rawls's commentary on *Methods* explicitly motivates the Kantian approach to morality that one finds in *KCMT*.[15] Although Rawls considers Sidgwick's *Methods* 'the outstanding achievement in modern moral theory,' he is not uncritical (*KCMT*, 554). In particular, he criticizes Sidgwick for misunderstanding and, consequently, underestimating a distinctly Kantian approach to moral inquiry (Rawls 1974–75, 9–10; *KCMT*, 556). In light of this criticism, it is natural to read his own work in *KCMT* as a corrective or supplement to Sidgwick's project in *Methods*. But, in order to understand how exactly *KCMT* is meant to correct or supplement *Methods*, we first need to know what Rawls means by 'moral theory' and how this contrasts with a different kind of project that he refers to by the name 'moral philosophy'. As we will see, the move from Sidgwickean to Rawlsian moral theory

---

[14]For an exception, see Terence Irwin's (2009) discussion of Rawls, esp., pp. 892, 897–98, 940, 945.

[15]See, e.g., Rawls (1974–75, 9–10); *KCMT* (554–56). Rawls also cites Sidgwick in his dissertation when first introducing his conception of moral theory (there under of the guise of 'traditional ethics'); see Rawls (1950, 5n).

reveals the particular way in which metaethical considerations enter into Rawls's characterization of Kantian constructivism. What's more, Rawls's treatment of these matters is controversial and does not map easily onto contemporary discussions of the so-called 'autonomy of ethics.' This is likely a further reason for why the details of Rawls's view in *KCMT* have received so little explicit attention.

Moral theory, according to Rawls, is the comparative study of substantive moral conceptions, the different moral structures they form, and the way in which these structures relate to moral sensibilities (or what would now more commonly go by the name 'moral psychology'). *Moral philosophy*, by contrast, is a broader notion that includes moral theory but also metaethical concerns in epistemology, metaphysics, the philosophy of mind, and the philosophy of language.[16] When Rawls writes of Sidgwick's *Methods* representing a model for *KCMT*, he is clearly referring to the former approach to moral inquiry. As the title of the lectures and subsequent commentary makes clear, Rawls understands his discussion of Kantian constructivism as contributing to a project *in* moral theory, and not one that is, in the first instance, in moral epistemology, metaphysics, or the philosophy of mind or language.

Some philosophers have taken Rawls's claim about 'the independence of moral theory' to mean complete neutrality concerning broader philosophical issues in, e.g., epistemology and metaphysics.[17] On such a view, moral theory is an approach in first-order ethics that has no bearing on one's second-order commitments; it is compatible with any number of second-order interpretations.[18] If this is in fact what Rawls means by 'inde-

pendence', the Conventional Interpretation must be mistaken. So it is important that I address this view at the outset.

A close reading of the passages in which Rawls makes these statements reveals that complete neutrality is not what Rawls has in mind. Rather, he intends 'the independence of moral theory' to express two different claims. The first is that moral theory is not methodologically subordinate to other branches of philosophical inquiry (i.e., metaphysics, epistemology, philosophy of language, and philosophy of mind)—instead, it stands in relations of mutual dependence with these areas of philosophy (Rawls 1974–75, 5–7, 21–22). Here Rawls is responding to a philosophical practice, one commonplace at the time he started making this claim about independence, that assigned methodological priority to questions in philosophy of language and epistemology and, secondarily, to questions in metaphysics and philosophy of mind. According to this approach to moral philosophy, one had to first settle questions, e.g., about the metaphysics of freewill or personal identity before addressing questions about basic moral principles and their relation to moral psychology. Rawls rejects this approach, but not on the grounds of neutrality ('To be sure, no part of philosophy is isolated from the rest', Rawls 1974–75, 6). Rather, he claims that the dependence relations, if any, run in the opposite direction ('Thus the problems of moral philosophy that tie in with the theory of meaning and epistemology, metaphysics and the philosophy of mind, must call upon moral theory', Rawls 1974–75, 6).

The second claim expressed is that we ought to bracket the notion of 'objective moral truth' until we have a firmer grasp of the results of moral theory and, specifically, how rival conceptions of moral sensibility compare.[19] Here the relevant distinction is one

---

[16]See Rawls (1974–75, 5–6); compare *KCMT* (554). As I will discuss in section 5, an early presentation of this distinction can already be found in Rawls's doctoral dissertation: see Rawls (1950, 136).

[17]For example, see Brink (1989, 306–07). It would also appear as if Freeman (2007, 233–34) takes this line.

[18]In more Kantian terms, one might describe this as the independence of moral theory from the commitments of theoretical reason (Freeman 2007, 234).

[19]See Rawls (1974–75, 7, 9). This explains why Rawls claims at the end of the lectures that '[i]t has been [his] intention to describe constructivism by contrast and not to defend it, much less to argue that rational intuitionism is mistaken' (*KCMT*, 570). One might wonder why Rawls would devote three lectures, spanning fifty-six pages of text, to Kantian constructivism if he were

between explication and justification.[20] Whereas moral philosophy is primarily concerned with how we ought to understand justification in the case of moral judgments (e.g., epistemologically or practically), moral theory is primarily concerned with *explication*—i.e., articulating and making explicit the candidate moral conceptions (including principles and structures) embedded in commonsense moral judgments and laying them out for comparison. Such an approach does not rule out the relevance of metaethical considerations (or questions of truth and metaphysics) but talk of 'metaethical commitments' is misplaced in this context. Instead, it is better to say that moral theory *involves* or *appeals to* metaethical views as a kind of practical postulate, one which explains and motivates a possible way of relating to oneself and others. As Rawls says, '[a]nd though doing this may involve settling theoretical difficulties, the practical social task is primary' (*KCMT*, 19).

In addition to revealing how *not* to interpret Rawls on the independence of moral theory, this discussion should also make clear that Rawls's disagreement with Sidgwick does not turn on an appeal to metaethical considerations. While it is true that *Methods* is primarily a work in first-order, normative theory, Sidgwick does not sharply distinguish the comparative philosophical project he pursues there from concerns in (what is now considered) metaethics.[21] In particular, he characterizes intuitionism in various ways—some of which make the view appear first-order, others more metaethical.[22] But this alone is perfectly

compatible with Rawls's own views about the independence of moral theory. If anything, then, the disagreement would have to concern the direction of dependence between these various orders of claims. As I will argue shortly, Rawls's critique of Sidgwick is that he is working with an impoverished conception of moral theory, one that forecloses the possibility that moral theory is methodologically prior to epistemology and metaphysics.

Rawls claims that Sidgwick failed to fully appreciate Kant's view as an independent rival to the moral conceptions he considers (i.e., egoism, intuitionism, and utilitarianism). But, as we will see, Rawls's correction does not merely involve adding Kantian constructivism to the list. Rather, Rawls thinks that Sidgwick failed to appreciate Kant's position because his approach in moral theory was not sufficiently detailed or rich to capture this possibility. So, in order to show that Kantian constructivism should be included, Rawls must first provide a more detailed account of what moral theory is and what elements it includes. In other words, we must advance beyond Sidgwickean moral theory to Rawlsian moral theory.

As indicated by the title of his work, Sidgwick individuates alternatives within moral theory by *methods*. According to Sidgwick, a *method of ethics* is 'any rational procedure by which we determine what individual beings "ought"—or what it is "right" for them—to do, or to seek to realise by voluntary action' (Sidgwick 1907/1966, 1). Rawls interprets Sidgwick to mean that these methods are 'specified by first principles . . . [that] aim at reaching true judgments that hold for all rational minds' (*KCMT*, 555).

> But a consequence of starting with methods of ethics defined as methods that seek truth is not only that it interprets justification as an epistemological problem, but also that it is likely to restrict attention to the first principles of moral conceptions and how they can be known. First principles are however only one element of a moral conception; of equal importance are its conception of the person and its view of the social role of morality. Until these other

---

not interested in defending the view—or he were not convinced that the view was preferable to rational intuitionism. The reason is that his discussion in the lectures is to be understood as a provisional stage in a larger comparative project in moral theory.

[20]Note that this distinction is first introduced in Rawls's dissertation (see 1950, 29) and is crucial to the structure of that work.

[21]Some, of course, would argue that the distinction between metaethics and substantive moral theory is anachronistic prior to the publication of Moore (1903).

[22]Compare, e.g., p. 96 with pp. 386–89, 496 of Sidgwick (1907/1966).

elements are clearly recognized, the ingredients of a constructivist doctrine are not at hand. (*KCMT*, 555)

The first correction suggested here is to characterize moral conceptions as more than mere methods (in Sidgwick's sense), where again this is primarily 'specified by first principles.' An explication of a candidate moral conception should also specify its respective conception of the person and its view of the social role of morality. Together, these three elements combine to form different moral conceptions for explication and comparison by way of moral theory.

A *conception of a person* is a moral ideal that represents a certain way of regarding oneself and one's capacities in relation to a more general description. Rawls refers to this more general description as the *concept of a person*.[23] This he defines as 'a human being capable of taking full part in social cooperation, honoring its ties and relationships over a complete life' (*KCMT*, 571). Thus, we might say that a conception of a person represents a particular specification of this concept of the person, one that involves a certain way of regarding oneself and one's capacities as 'the basic unit of agency and responsibility in social life' (*KCMT*, 571). Rawls allows that there will be various such conceptions and that each of which may be further individuated in response to the social role it is supposed to play.

The *social role* of morality describes a particular task or problem that a specific conception of a person is supposed to respond to or solve (*KCMT*, 571). As Rawls notes, this role can either take a narrow or wide scope, depending on the extent to which the conception is supposed to function as a regulative ideal. According to the narrow scope, the task of a self-conception involves 'achieving more or less the minimal conditions of effective social cooperation' (*KCMT*, 553). Here, morality functions as a minimal framework of side-constraints, one that governs a par-

ticular domain of social interaction but which, otherwise, leaves us largely free to determine and pursue other ends. In other words, it does not inform or provide the justificatory grounds for everything we do. According to the wide scope, by contrast, the task of a self-conception and its related moral principles is to effectively govern all of our deliberations and actions. In the words of J. L. Mackie, it informs an 'all-inclusive theory of conduct.'[24]

Rawls thinks that once these additional two elements are incorporated into our framework for moral theory we will not only be in a position to see the logical space for new and unappreciated candidates in moral theory, ones that Sidgwick neglects (like constructivism and perfectionism), but we will also come to see that the familiar candidates (like egoism, utilitarianism, and intuitionism) take on new dimensions. A full description of each of these moral conceptions will include a set of first principles, a conception of a person, and an account of the social role these elements are supposed to play.

The second correction that Rawls makes involves allowing for the possibility that moral personality and the social role of morality take priority over, and determine the content of, first principles. Once we have expanded our framework in moral theory to include more than one element, we are able to pose the question as to how these elements are supposed to relate to one another. There is the possibility that some elements will take priority over others. Not every moral conception will give primacy to the conception of the person and its social role, but some will; others will continue to assign primacy to first principles. In contrasting these alternatives, Rawls makes explicit appeal to metaethical considerations—albeit from within the engaged, first-order perspective of moral theory.

For example, rational intuitionism may be described in terms of first principles, a conception of the person, and a statement

---

[23]According to Paul Weithman, Rawls takes the concept/conception distinction from H. L. A. Hart. See Weithman (2010, 32).

[24]See Mackie (1977, 106). Note that Rawls specifically cites Mackie as the source of this distinction (*KCMT*, 553).

of the social role of morality. But, on this view, as the truth of principles is independent of and prior to moral personality, it only requires 'a sparse notion of the person, founded on the self as knower... the only requirements on the self are to be able to know what these principles are and to be moved by this knowledge' (*KCMT*, 560). To the extent that rational intuitionism is understood as a view in moral theory, and not just a view in moral epistemology and metaphysics, it involves this self-conception taking a wide social role. In other words, one is to entertain what it would be like to regard oneself *as if* there were stance-independent moral truths that we come to know via a faculty of rational intuition and then ask how this form of self-regard would structure norms for relating to ourselves and others.

Importantly, Rawls appears to believe that conceiving of oneself under the intuitionist's description does affect the form of one's self-regard in ways that have practical consequences. By contrast, Rawls is keen to stress that constructivist moral conceptions would have us regard ourselves very differently and that this difference matters from within the standpoint of moral theory. We are now in a position to appreciate the way in which Kantian constructivism both compares with rival theories and contributes to our understanding of the structure of and possibilities for an approach to moral theorizing that assigns primacy to conceptions of moral personality. However, before presenting this constructivist alternative, it is worth noting the particular way in which Rawls appeals to metaethical considerations in moral theory and the bearing he thinks these have on one's broader theoretical commitments.

Strictly speaking, one may accept an intuitionist (or a constructivist or a perfectionist) view within moral theory without theoretically committing oneself to the associated views in metaphysics and epistemology. But to do so would appear to invite an error theory or, at the very least, present a challenge for achieving wide reflective equilibrium amongst one's beliefs. Although Rawlsian moral theory does not assign primary focus to metaethical considerations, it does in part individuate alternative candidate views according to their metaethical features and evaluate them on this basis. To the extent that a package of metaethical views lends itself to an attractive self-conception (one that finds support in our deeply-held moral and evaluative judgments), this will count in favor of our accepting such views both from within moral theory and, as a further consequence, from the perspective of moral philosophy, more broadly.[25] Again, Rawls's approach to moral theory allows that the dependence relations run in this direction; he only disputes the methodological subordination of moral theory to metaphysics, epistemology, and philosophy of language. Rawls's aim is not to quarantine metaphysically robust ways of talking and thinking about morality to an engaged, first-order perspective. His point is, rather, that these ways of talking and thinking have different implications for our self-conceptions and that the plausibility or attractiveness of these self-conceptions matters not only for how we relate to ourselves and others but also what views we ought to accept in moral philosophy more broadly.

4. Kantian Autonomy and the Primacy of Moral Personality

Kantian constructivism is a view in moral theory that assigns priority to a Kantian conception of moral personality and autonomy. According to such a self-conception, we are to regard ourselves as free and equal persons whose actions are only normatively constrained by the demands of our own practical reasoning. On this way of conceiving ourselves, there is no in-

---

[25]Again, this reveals the way in which Rawlsian moral theory forms part of a project of internal accommodation (compare Timmons 1999, 11–12).

dependent or higher authority than that which we license our-selves. This is what Rawls means when he says that we are to regard ourselves as 'self-originating sources of valid claims' (*KCMT*, 543). Here, by contrast with intuitionism, we are to sup-pose a very different view in moral metaphysics. In this case, the Kantian self-conception Rawls describes is hardly intelligible without the supposition that a robust, stance-independent real-ism is mistaken. This is perhaps most evident in the way Rawls contrasts Kantian autonomy with the heteronomy of rational intuitionism.

Rawls's historical work on Kant illustrates this point espe-cially well. In his 'Themes in Kant's Moral Philosophy' (1989), Rawls summarizes *rational intuitionism* as the combination of three claims:[26] first, basic moral concepts are not analyzable in terms of non-moral concepts; second, moral first principles are 'true statements' that serve to justify claims about what is good, what is right, and what has moral worth; third, 'first principles, as statements about good reasons, are regarded as true or false in virtue of a moral order of values that is prior to and independent of our conceptions of person and society, and of the public social role of moral doctrines' (Rawls 1989, 95). Although Rawls states that the first two claims are held ' in common with a number of other views, including Kant's' (Rawls 1989, 95), he argues that the third claim is rejected by Kant. On Rawls's interpretation, Kant must reject the third claim on the grounds of heteronomy.

Rawls argues that rational intuitionism is heteronomous and that, as such, it cannot be Kant's view. As he explains, it is het-eronomous because it assigns priority to an independent order of moral facts, but Kantian moral autonomy, on his account, 're-quires that there exists no moral order prior to and independent of those conceptions that is to determine the form of the pro-

cedure that specifies the content of first principles of right and justice among free and equal persons' (*KCMT*, 512). We see that much hangs on how Rawls understands the priority relation. Although Rawls sketches different approaches to understand-ing this relation, he appears to favor an *explanatory* notion.

> Another way to state the relation of priority is to say that it concerns the order of explanation. Rational intuitionism says: the procedure is correct because following it correctly usually gives the correct (independently given) result. Constructivism says: the result is correct because it issues from the correct reasonable and rational procedure correctly followed. (Rawls 2000, 242)

Although Rawls talks here in terms of results being 'correct', it is important to note (for reasons that will become apparent later) that Rawls does not say that Kant rejects rational intuitionism because he rejects the view that moral statements are *true or false* or, alternatively, the view that moral truth is correspondence with the facts. Indeed, that he also characterizes rational intu-itionism in terms of correctness in this passage, a view that he undoubtedly understands in terms of truth claims, suggests that we should not read too much into the specific terms he employs here. According to Rawls, Kant would take rational intuition-ism to be heteronymous because it gives the wrong account of why moral statements are correct or incorrect—and, from what he says elsewhere, it turns out that that this is just to say that it gives the wrong account of why moral statements are true or false.

It would appear then as if Rawls's interpretation of Kant's constructivism is captured by the Conventional Interpretation I started with. According to this interpretation, Kant maintains that there are true first principles of morality and that what *explains* their truth is that they result 'from the correct reasonable and rational procedure correctly followed.' This is just to restate the earlier characterization of constructivism as appealing at some level to a cognitivist, stance-dependent, success theory.

---

[26]See Rawls (1989, 95–96). Something like this account is repeated in (1993/1996, 90–92). However, there he states the view in four points. At *KCMT* (557), rational intuitionism is presented in two points.

Rawls, of course, explicitly denies that the constructivism in *KCMT* is entirely modeled on Kant's view and flags specific differences between the two.[27] For example, whereas Kant's account of the *Categorical Imperative* applies to the personal maxims of individuals in everyday life, justice as fairness assigns primacy to the social—i.e., to unanimous collective agreement (*KCMT*, 552–53). However, these points of departure from the historical Kant do not appear to affect the claim that the constructivism of *KCMT* involves appeal to a stance-dependent metaethical doctrine in its conception of the person; rather, this specific difference merely concerns the description of the choice procedure that is to determine the content of moral principles. In fact, if we look at those passages in *KCMT* where Rawls contrasts his own view with rational intuitionism, we see that the contrast is presented in the very same terms.[28] There is no recognizable difference between the way that Rawls interprets the historical Kant on this point in later works—e.g., his (1989) and (2000)— and the way that he presents his own constructivism in *KCMT*. The only possible difference might concern the scope of Rawls's constructivism in the lectures. Whereas, in *KCMT*, Rawls is primarily concerned with positing a constructivist conception of moral personality and working out the further consequences this would have within moral theory, it is not clear that Rawls takes Kant to be working within a similarly restricted framework.

---

[27] Rawls prefers to call his own view 'Kantian', a designation that he claims 'expresses analogy and not identity' (*KCMT*, 517). For the same point, see Rawls (1985, 224n). For specific differences between Rawls's view and Kant's, see (*KCMT*, 552, 566; Rawls 1971/1999, 226–27). In later works, the differences begin to make Rawls's view look much less like a metaethical view. See Rawls (1985, 247; 1993/1996, 89–129, esp. 99–101).

[28] Compare the above quoted passage from *KCMT*, 568. Note also the degree of similarity between Rawls (1989, 95–102) and *KCMT* (557–60)—which includes the verbatim repetition of some passages.

5. Responding to Interpretive Challenges

But even if Rawls appeals to metaethical considerations from within moral theory in the way that I have argued, it remains to be seen whether the specific metaethical interpretation appealed to in the constructivist self-conception is best characterized in terms of a stance-dependent cognitivism. In particular, one might still wonder how Rawls could coherently characterize constructivism in these terms if he also disavows truth and metaphysics. The key to understanding Rawls's apparent disavowal of truth and metaphysics in *KCMT* does not lie in his comments on the independence of moral theory or in *A Theory of Justice* (hereafter *TJ*) or even in *Political Liberalism* (hereafter *PL*) but, rather, in Rawls's earliest work, viz., his doctoral dissertation. As I shall argue, Rawls's remarks in the dissertation allow us to reconstruct a view of moral truth that both positively supports the interpretation of Kantian constructivism that I have presented and provides reason to discount rival interpretations, including those that would have us read the constructivism of the lectures as incompatible with cognitivism, stance-dependence, or truth as correspondence. Before presenting the case, however, I will first explain why we should not expect to find help in familiar places.

5.1. Sources

Unfortunately, Rawls's comments on the independence of moral theory do nothing to support the *specific* metaethical picture that one finds expressed in the Conventional Interpretation. Recall that this characterization posits a kind of stance-dependent cognitivism, one according to which moral statements are made true by facts about what free and equal persons would agree to. But one could characterize Kantian constructivism in metaethical terms and remain independent in the relevant sense(s) and yet deny stance-dependence, or cognitivism, or truth as correspon-

dence. Furthermore, while there may be nothing about moral theory *per se* that precludes appeal to truth and metaphysics, Rawls might appeal to other, independent considerations that, on the whole, provide reason to remain quiet on these issues.[29] So, we must look elsewhere for an answer to whether Rawls's apparent disavowal of truth and metaphysics in *KCMT* is compatible with a characterization of constructivist moral personality that involves a stance-dependent cognitivism.

A response to the interpretative challenge is also not to be found in either *TJ* or *PL*, despite the similarities between these works and the lectures. Although there is much continuity between *KCMT* and *PL*, it is now commonly acknowledged that Rawls's characterization of political constructivism in the later work represents a change in his views and not merely a clarification of the earlier statement (despite what Rawls's comments would suggest[30]). In what follows, I present evidence of one substantive change in Rawls's presentation which I believe sufficiently illustrates the differences between the views in these texts.

The argument Rawls makes in *PL* for favoring reasonableness over truth may apply to *political* constructivism, but it does not extend to Kantian constructivism more generally. In *PL*, Rawls's disavowal of truth is clearly motivated by the importance of *public reason* for his conception of justice as fairness (see Rawls 1993/1996, xxii, 94, 116, 126–29, 153). Rawls argues that this conception requires citizens to restrict themselves to certain types of

premises and forms of reasoning when making public political arguments. These premises and forms of reasoning, he says, must provide a public basis that can support an overlapping consensus of reasonable political views. Otherwise, they could not serve as a basis for the kind of agreement that can ensure stability for the right reasons in a pluralistic democratic society.

Rawls argues that political arguments based on the commitments of any one group's comprehensive system of moral, philosophical, or religious beliefs would fail to satisfy the requirements of public reason and, consequently, would not provide the right kind of justification for the terms of social cooperation. In particular, Rawls worries that an appeal to the truth of one's moral or political beliefs would be unnecessarily divisive and thinks that a public basis of justification acceptable to all is more likely to be achieved by restricting our claims to what is reasonable (Rawls 1993/1996, xxii, 94, 116, 126ff, 153–54): 'while people can recognize everyone else's comprehensive views as reasonable, they cannot recognize them all as true, and there is no shared public basis to distinguish the true beliefs from the false.'[31] Whereas Rawls's claims about truth and public reason are controversial, it is not controversial that these claims explain Rawls's preference for 'reasonable for us' over 'true' in later texts, like *PL*.[32] But whatever we think of Rawls's argument in *PL*, it cannot be the same argument that Rawls appeals to in *KCMT*.

Recall that, in *KCMT*, Rawls states that justice as fairness only represents one particular variant of Kantian constructivism in moral theory. In other words, Kantian constructivism is a gen-

---

[29]In fact, as I will argue shortly, this is exactly what Rawls does in later work (compare 1993/1996, xxii, 94, 116, 126–29, 15).

[30]In the introduction to *PL*, Rawls claims that 'the first three lectures more or less cover the ground' of the three lectures which comprise *KCMT* and that, in their latest revised form, he thinks that 'they are much clearer than before' (1993/1996, xv). In Rawls (1985, 224n), he notes that the earlier work was 'misleading'. Other similarities between the lectures and *PL* notwithstanding, these statements do not accurately characterize the relation between Rawls's presentation of constructivism across the two works.

[31]Rawls (1993/1996, 128). Rawls voices similar concerns in *KCMT* (541–42). But there his response is not to disavow truth altogether; rather it is to restrict justification in his preferred choice procedure (i.e., *the original position*) to a part of the truth, instead of the whole of truth.

[32]For example, Cohen (2009, esp. 6ff) agrees that Rawls's disavowal of truth is grounded in concerns about the importance of public reason, but he argues that Rawls is mistaken to think that there is no place for appeals to truth in public reason.

eral view in moral theory that includes both political and non-political moral notions. Rawls, however, allows that the non-political moral notions are unencumbered by the constraints of public reason. In fact, he stresses that his discussion of public reason 'applies only to the principles of political and social justice and not to all moral notions' (*KCMT*, 539; see also 538). If this is correct, Rawls's argument from *PL* for preferring reasonableness to truth does not extend to these other moral notions. Furthermore, Rawls is clear that a disavowal of truth in favor of reasonableness characterizes Kantian constructivism about these non-political moral notions, as well.[33] Taking these two points together, one must conclude that Rawls's argument for disavowing truth in *KCMT* cannot be the same as the one he offers in *PL*.[34]

---

[33]The broader characterization of constructivism is clearly operative in several key passages in which Rawls contrasts the constructivist criterion of reasonableness with the criterion of truth—as represented both in Sidgwick's work (*KCMT*, 555) and rational intuitionism more generally (557).

[34]Here, it is also important not to confuse the different senses in which Rawls's views in *PL* and *KCMT* are *neutral* with respect to a theory of truth. In *PL*, political constructivism is neutral in the sense that one may hold whatever view of truth one wants (assuming one has views on this) so long as one does not appeal to it in framing and advancing political arguments. In other words, a theory of truth has no direct role to play in political justification. In *KCMT*, by contrast, a theory of truth may play a foundational role in characterizing and individuating rival candidate self-conceptions within moral theory and, consequently, determining the content of first principles. But Rawls's characterization of Kantian constructivism in *KCMT* is underdetermined on this point, and so, in a different sense, might be said to be neutral with respect to a theory of truth. As I will argue later, what Rawls says is perfectly compatible with a correspondence theory of truth and an ontology of truth-makers. But, on its face, it is similarly compatible with minimalism or otherwise deflationary views of truth. Hence, in principle, we may speak of different species of Kantian constructivism depending on the theory of truth they incorporate at the level of self-conception. I say *in principle*, since it might turn out that this difference is of no practical import. One of the tasks of moral theory would be to compare these rival self-conceptions and work out the extent to which, if any, this difference with respect to truth would generate different forms of self-regard and/or regard for others.

One might pivot from this conclusion to the thought that the key to understanding *KCMT* on these points must then lie in *TJ*. But this would also be a mistake. There are important differences in the presentation of the 'Kantian Interpretation' of *TJ* and the constructivism of *KCMT*. One such difference involves the way that Rawls presents intuitionism in each work.

Again, in the lectures, Kantian constructivism is presented as a competitor to *rational* intuitionism, a view that includes metaethical elements and according to which first principles are 'fixed by an order of universals or concepts' (*KCMT*, 559), one 'given by the nature of things and . . . known, not by sense, but by rational intuition' (*KCMT*, 557). Importantly, Rawls describes this view as 'compatible with a variety of contents for the first principles,' including, e.g., utilitarianism, perfectionism, or pluralism (*KCMT*, pp. 557, 559). In *TJ*, by contrast, he characterizes a view he calls 'intuitionism' as a first-order competitor to both justice as fairness and classical utilitarianism, one according which there is a 'plurality of first principles . . . and . . . no explicit method, no priority rules, for weighing these principles against one another' (Rawls 1971/1999, 30). Although Rawls acknowledges that this form of pluralism has traditionally been packaged with epistemological commitments, he states that these are 'not a necessary part of intuitionism as I understand it' (31). Fittingly, then, the Kantian Interpretation in *TJ* is presented as an alternative to this first-order view. Although the Kantian ideal of a person appealed to there articulates a set of 'constructive criteria' that support Rawls's favored priority rules, the view should not be confused with Kantian constructivism. For this reason, we should not expect to find help in *TJ* for interpreting the view Rawls later presents in *KCMT*. But interpretive help can be found elsewhere, albeit in a less familiar place.

Although Rawls does not appear to have appreciated the possibility of Kantian constructivism in his dissertation, one can find many of the same elements there that one finds in *KCMT*

and other works in which he is primarily focused on moral theory. Indeed, the degree to which the dissertation previews and sets the stage for these later works is remarkable. Despite some important differences in the details of the substantive views Rawls entertains, there is considerable overlap and continuity in the methodological discussion. For example, Rawls distinguishes the project in moral theory (which goes by the name 'ethical theory' in this context) from other areas of philosophical inquiry in more or less the same terms we find in these later works (Rawls 1950, 136). He also introduces the distinction between truth and reasonableness in his dissertation.[35] But that is not all. Most importantly, he is also far more explicit there about how he understands this distinction and why he introduces it (Rawls 1950, 247).

5.2. Kantian Constructivism as a Cognitivist Self-Conception

Rawls provides a very helpful and concise explanation of his views on moral truth in a section of the dissertation titled, 'Are Ethical Judgments true or false:'

> Ethical judgments, as assertions that certain maxims are justifiable, are plainly either true or false. The maxim in question may be, or may not be, a case of, or an instance of a justifiable principle. Yet the maxim itself is not properly said to be true or false, but preferably, reasonable or unreasonable, justifiable or unjustifiable. The property which the justifying process decides with regard to a maxim is not that of being true, or that of being false, but that of being reasonable or unreasonable. It is only the assertion (the judgment) that it has, or has not, this property, which is true or false. (Rawls 1950, 255–56)

There are two important points to take away from this passage. The first is that maxims (and the principles which they are in-

stances of) are not the bearers of truth; rather ethical judgments, as assertions of claims (or sentences or propositions, etc.), are. Strictly speaking, then, we should say that maxims or principles are *reasonable* (or *unreasonable*), despite its being perfectly acceptable to say that the claims (or sentences or propositions, etc.) that express these maxims or principles are true (or false).[36] The second is that, on Rawls's view, ethical judgments do not attribute any special kind of property to things, or states-of-affairs; rather, as he makes clear elsewhere in the dissertation, they assert that a maxim is justified in virtue of its 'logical' relation to a principle (Rawls 1950, 280). Unlike the kind of non-naturalist ethical realism that Rawls explicitly rejects, the view he presents here does not involve positing any special class of ethical properties, one that would involve controversial ontological commitments (Rawls 1950, 260–64, 280).

These earliest expressed views on moral truth are compatible with what Rawls says in the lectures. To the extent that Rawls's account changes at all in *KCMT*, this involves a broadening (not a narrowing) of the scope for moral truth. For example, when Rawls explains his substitution of 'reasonable' for 'true' in the lectures, he states that it is still appropriate to speak of moral truths at the level of particular judgments and inferred principles.

> This usage, however, does not imply that there are no natural uses for the notion of truth in moral reasoning. To the contrary, for example, particular judgments and secondary norms may be considered true when they follow from, or are sound applications of, reasonable first principles. These first principles may be said to be true in the sense that they would be agreed to if the parties in the original position were provided with all the relevant true beliefs. (*KCMT*, 569)

Here, Rawls allows that we may speak of truth not only as a property of moral judgments (or assertions) but also of first prin-

---

[35]The first occurrence of this distinction can be found in Rawls's discussion of the difference between scientific and ethical knowledge. See Rawls (1950, 100).

[36]Throughout the dissertation, Rawls is in fact quite liberal in his use of 'moral truth'. See Rawls (1950, 35–36, 88, 192, 264).

ciples 'in the sense that they would be agreed to' in the original position.[37] I take this passage (together with the dissertation, in which Rawls explicitly rejects non-cognitivist approaches to ethical inquiry: Rawls 1950, 1–3, esp. 3n) as sufficient evidence for rejecting any kind of non-cognitivist reading of the constructivism of the lectures. The fact that Rawls recognizes some natural uses of the notion of moral truth indicates that his use of 'reasonable' and 'unreasonable' for moral discourse does not involve rejecting cognitivism; rather, it suggests that 'reasonable' must pick out some other difference.[38]

We get a good sense of what this difference is supposed to be by examining Rawls's claim that truth cannot be the *aim* of the parties to the construction procedure and that 'the idea of approximating truth has no place in a constructivist doctrine' (*KCMT*, 564). In the original position, the parties do not aim at some independent notion of moral truth that their decisions might approximate because, by hypothesis, prior to the procedure 'there are no such moral facts to which the principles adopted could approximate' (*KCMT*, 564). According to constructivism, the procedure *determines* what the moral facts are by first arriving at a reasonable set of principles; 'the facts are [then] identified by the principles that result' (*KCMT*, 568). There is no

external or procedurally-independent criterion that can be appealed to. Hence, the procedure is an essentially ineliminable feature of the constructivist's account of morality. This is what Rawls means when he calls justice as fairness, the variant of constructivism he is most concerned with, a form of 'pure procedural justice' (*KCMT*, 522). If the procedure did not determine the moral truths—i.e., the truth-makers for moral statements—but merely reliably detected or approximated them, it would be eliminable; hence, there would be no distinctly constructivist position to speak of.[39] So, if, as the Conventional Interpretation would have it, agreement in the original position determines what the moral truths are, the parties to the choice procedure must have a different aim than truth. This is why Rawls describes their aim in terms of reasonable agreement.

By contrast, from the standpoint of moral agents outside of the original position deliberating about what to do in the everyday, the outcomes of the procedure constitute the standards in virtue of which actions count as required, forbidden, or permissible.[40] In other words, the particular judgments agents make are made true, or false, depending on the more general principles that would result from the choice procedure. This is not to say that the main point or function of reasonable first principles is to underwrite the truth of ethical judgments. Rather, the search for such principles as the grounds for our actions is a practical pursuit, one whose aim or function is to make possible a form of life characterized by free and equal moral personality taking a wide social role. Nevertheless, this view is compatible with ethical cognitivism; indeed, this is clearly the framework that Rawls himself appears to be working with in *KMCT*.[41]

---

[37]This would appear to reflect one important difference between the constructivism of *KCMT* and the view Rawls endorses in the dissertation: see Rawls (1950, 284–85). There he is clear that satisfying the (final) test for the reasonableness of a maxim is not the same thing as being reasonable. Rather, the former is merely evidence of the latter. In other words: *reasonableness* is not reducible to a stance-dependent property.

[38]Of course, defenders of more sophisticated versions of non-cognitivism, like expressivism, argue that a non-cognitivist can 'earn the right' to take on realist ways of thinking and speaking about morality, including notions of moral knowledge, truth, and objectivity. Part of the expressivist strategy involves taking on board a *minimalist* view of truth, e.g., one according to which truth is not a substantive property. See Blackburn (1998, 77–83); Gibbard (2003, 18). But, as I will later show, there is also good reason for rejecting an interpretation of *KCMT* that requires a non-correspondence account of truth.

[39]For an extended discussion of the claim that the constructivist procedure is an essentially ineliminable feature of constructivist views, see Enoch (2009, 331ff). See also Barry (1989, 264–82).

[40]See *KCMT* (569). Note that Rawls himself emphasizes the importance of distinguishing these different 'points of view' (*KCMT*, 533–34; 567–70).

[41]I thank one of my readers for asking me to clarify this point.

5.3. Kantian Constructivism as Stance-dependent Cognitivism

But even if one accepts that the Kantian constructivism of the lectures involves an appeal to a form of cognitivism, one might still doubt other metaethical features attributed by the Conventional Interpretation. For instance, one might worry that a pure procedural conception of justice and morality does not count as a *stance-dependent* form of cognitivism. Stance-dependence would appear to be neither necessary nor sufficient for pure proceduralism.

For example, on a desire satisfaction theory of well-being, being *good for* someone is a stance-dependent property—one 'that exists only if some thing or state of affairs is made the object of an intentional psychological state (a stance),' viz., the agent's desire(s). But such a theory of well-being need not appeal to procedures.[42] So stance-dependence alone is clearly not sufficient for pure proceduralism.

It is also not necessary. For example, we might imagine pure procedural justice in a particular context being realized by a series of fair coin flips. In this case, we may suppose that the outcome of the procedure is fair regardless of whether it is one that the parties would desire or choose or agree to. Moreover, even procedures that involve agreement or choice need not appeal to stance-dependence as an essential feature. Compare, for example, an actual legislative body that deliberates and makes laws. Even if one grants that the legislative process 'models' or 'represents' a particular deliberative standpoint, the validity of the legislative outcomes need not depend in an essential way on the actual psychological states (or stances) of the par-

ties to the agreement. Rather, this validity may merely depend on the satisfaction of superficial procedural conditions (e.g., the parties' exhibiting behavior that satisfies certain descriptions), not on anything 'in their heads' (i.e., their particular beliefs, desires, or expressions of will). So the pure procedural approach characterized above would appear to be neither necessary nor sufficient for the stance-dependence attributed by the Conventional Interpretation and that I claim is appealed to as part of the constructivist self-conception in moral theory.

Although there is nothing about pure proceduralism *per se* that requires or is entailed by stance dependence, the specific form of hypothetical proceduralism that Rawls describes in the lectures does, nevertheless, count as a form of stance-dependent cognitivism. The original position 'incorporates pure procedural justice at the highest level' (*KCMT*, 523), but pure procedural justice alone does not amount to constructivism (compare Barry 1989, 266). Rather, we only have constructivism when the procedure models, or represents, a particular conception of the person in a way that generates (i.e., *constructs*) the content of first principles.[43] In the context of the original position, this involves specifying the principles that the more general and abstract moral ideal of free and equal personality taking a wide social role would commit one to. Nevertheless, constructivism is not reducible to a form of coherentism, one that merely involves bringing our antecedent beliefs about freedom and equality into reflective equilibrium.

Whereas 'general and wide reflective equilibrium' plays an important role when we are at the final stage of comparing and evaluating the candidate moral conceptions that we, as theorists, have explicated, it would be a mistake to think that the original position is merely a heuristic for bringing our intuitions at var-

---

[42]While it is true that sophisticated versions of a desire satisfaction theory might appeal to one, or another, form of idealization and this typically *does* take a procedural characterization, this is not the case for views that rely on actual or present desires. Thanks to an anonymous referee for helping me to clarify this point.

[43]Rawls calls the original position a 'mediating model conception'. By this, he means that it is 'a device of representation used to connect the conception of the person with definite principles of justice' (*KCMT*, 533–34).

ious levels of generality into reflective equilibrium. Reflective equilibrium is not directly appealed to within the original position as a way of specifying the content of first principles. Mere reflection on abstract moral ideals, like free and equal moral personality or full autonomy, is not sufficient for this task; rather, the theorist must construct, or generate, the content of first principles by working through the procedure hypothetically.[44] Because our hypothetical agents of construction are characterized in terms of the moral psychology they posses, we determine the content of first principles by imagining what persons with such sensibilities would commit themselves to, i.e., what volitional *stances* they would bear toward various possible sets of principles. It is for this reason that a constructivist form of pure proceduralism represents a *stance-dependent* cognitivism.

5.4. Kantian Constructivism and Truth as Correspondence

This brings us to the last worry that I will consider. According to this objection, one might grant stance-dependence and yet deny that Kantian constructivism appeals to an account of truth as correspondence with the facts. The main argument for thinking this is that, if Kantian constructivism included truth as correspondence, this would involve controversial ontological commitments. But Rawls explicitly rejects such commitments. So, Kantian constructivism should not involve any appeal to truth as correspondence. To be clear, I do not believe that the text requires that we characterize Kantian constructivism as appealing to truth as correspondence. The available textual evidence is underdetermined on this point and would appear compatible with a reading of constructivism that is agnostic on the question of truth or even readings that appeal to a minimalist, or

otherwise deflationary, conception of truth. However, as I will show, the available evidence is *also* compatible with a correspondence conception of truth; Rawls's worries about metaphysics are actually much narrower in their scope. In fact, I would argue that including such a characterization would better serve to accentuate the contrast that Rawls intends to draw between constructivism and rational intuitionism than rival conceptions would.

There are two plausible rival interpretations that I need to respond to here, both of which take the argument above as motivation for attributing a non-correspondence account of truth to the constructivism of *KCMT*. Each of these rivals would have us suppose that the nature of truth in morality is different from the nature of truth in other discourses and that Rawls's use of the term 'reasonable' is a way of marking this difference. Although both of these views are interesting and plausible, in the sense that they describe paths Rawls could have taken, I will show that they each fail to take account of the evidence and, in particular, the project in moral theory that Rawls is pursuing in the lectures.

According to the first, advanced by David Brink, the Rawls of *KCMT* maintains a coherentist theory of truth for morality, though not for other discourses.[45] On this view, the term 'reasonable' flags an ontologically minimal, *epistemic* conception of truth for morality—one that takes the truth of a moral statement to be what would be believed in reflective equilibrium, not correspondence with the facts. Brink argues that ideals of moral personality are underdetermined in a way that underdetermines theory choice in ethics (Brink 1989, 315). In particular, the ideal of free and equal moral personality admits of different interpretations; these different interpretations support differ-

---

[44]Barry (1989, 271–82) argues that, otherwise, constructivism would be indistinguishable from intuitionism, understood in a broad sense.

[45]Brink (1989, 305, 307, 310–12). Michael Lynch is sympathetic to this kind of view in general and cites Brink's interpretation of *KCMT* in particular. See Lynch (2009, 165ff).

ent sets of first principles. But, because the ideals themselves are insufficiently detailed to adjudicate disputes between their competing interpretations, we are also unable to adjudicate disagreements about first principles. Brink takes Rawls's solution to this problem to be to relativize moral truth to the beliefs, both moral and non-moral, that we possess—including 'those moral beliefs about persons on which such theories depend (evidentially)' (Brink 1989, 315, cf. 305). In other words, a coherentist theory of moral truth is a response to a metaphysical worry about indeterminacy at the level of ideals of moral personality.

Despite its attractions, Brink's rational reconstruction fails to make sense of the textual evidence in three ways. First, Rawls explicitly denies that his view in *KCMT* relies on a different theory of truth for morality.

> Furthermore, it is important to notice here that no assumptions have been made about a theory of truth. A constructivist view does not require an idealist or verificationist, as opposed to a realist, account of truth. (*KCMT*, 565–66)

I take this passage, together with others in the vicinity, to provide good evidence for rejecting the view that the nature of moral truth differs from truth in other discourses.[46] Moreover, it is natural to read the second sentence as saying that constructivism is, at the very least, compatible with a 'realist' account of truth, by which I take him to mean a correspondence theory.

Second, nowhere in the lectures does Rawls assert that constructivism is a response to metaphysical indeterminacy or that ideals of moral personality are underdetermined in a way that underdetermines theory choice in ethics. In fact, he regularly speaks of justice as fairness as presenting 'the most reasonable' set of principles for the practical task at hand (*KCMT*, 517, 519, 534, 541, 547, 554, 569–70). He also explicitly allows for the possibility 'that constructivism is compatible with there being, in fact, only one most reasonable conception of justice' (*KCMT*, 570) or a 'single most reasonable conception' (*KCMT*, 569). Moreover, considering that the constructivism of the lectures is presented within the framework of moral theory, we should not expect Rawls to take a stand on determinacy or objective truth.[47] Again, the second sense in which moral theory is independent is that it sets these kinds of questions aside until a later stage when we have a firmer grasp of how rival conceptions of moral sensibility compare (see Rawls 1974–75, 7, 9).

Third, and lastly, it is interesting to note that Brink's argument takes as its starting point the same observation I appealed to above: that the general and abstract ideals of moral personality that Rawls starts out with are insufficient for specifying determinate content for first principles. However, there I argued that this is only a problem if one undersells the role of the original position in generating content and, instead, takes Rawls to be assuming a kind of coherentism, one according to which the original position is a heuristic for bringing our intuitions at various levels of generality into reflective equilibrium. That Brink attributes coherentism to Rawls in the lectures suggests that he does not to fully appreciate what Rawls has in mind when he describes the original position as incorporating pure procedural justice at the highest level—in particular, that he intends the original position to specify and generate the content of

---

[46]Brink acknowledges this passage but argues that Rawls is only rejecting an epistemic ('antirealist') theory of truth for *non-moral* statements (1989, 310–11), hence, supporting his truth pluralist reading. Brink, however, does not take into account Rawls's restatement of these claims at *KCMT* (569). There, *pace* Brink, it is clear that Rawls intends his claims to apply to the nature of moral truth, as well.

[47]Here I mean a stand from within the broader theoretical perspective of moral philosophy. Although Rawls's account of Kantian constructivism is underdetermined with respect to the specific nature of truth (i.e., whether this characterization involves a correspondence theory or its rejection), there is nothing about moral theory *per se* that would count against individuating rival conceptions of moral personality in terms of their specific views about the nature of moral truth.

first principles in a way that goes beyond an appeal to reflective equilibrium.

According to a second interpretation, the constructivism of the lectures is to be understood as a form of cognitivism, in the sense of allowing for a kind of practical *knowledge*, but not one that is grounded in a theoretical conception of truth, correspondence or otherwise.[48] The advocates of this interpretation would object that, if we assumed the latter, such a view would fail to capture what is special about the Kantian distinction between practical and theoretical reason. This is where they would argue that the Conventional Interpretation (and, in turn, my appeal to such a view as part of the constructivist self-conception in moral theory) gets things wrong. It involves an *ontologically-based account of practical reason* at the level of secondary principles and moral judgments (Bagnoli 2013a). That is, at this level, it would characterize the task of practical reason in terms of identifying facts (albeit stance-dependent ones) and then applying this knowledge in practical deliberation. But here these opponents would argue that, to the extent that such an account relies on facts external to the activity of practical reason itself, it will be at odds with a Kantian account of the autonomy of practical reason and, consequently, fail to accommodate the bindingness and categorical authority of moral judgments (Korsgaard 2003). What's more, these opponents might claim that such a view would be at odds with Rawls's own rejection of methods that conceive of moral justification as an epistemological problem and not a practical one (*KCMT*, 554–55, see also 518–59). Hence, we should read Rawls's substitution of 'reasonable' for 'truth' as intended to capture this difference between practical and theo-

retical reason and their respective conceptions of objectivity and knowledge.

Although a more thoroughgoing constructivism about practical reason may turn out to represent a more promising way of capturing what is distinctive about the constructivist challenge to traditional approaches in metaethics, this is decidedly not the characterization of constructivism one finds in *KCMT*. While Rawls is clearly critical of methods that conceive of moral justification in purely epistemological terms and that, consequently, employ only a 'sparse notion of the person, founded on the self as knower' (*KCMT*, 560, see also 571), it is also clear that he intends these comments to be restricted to the determination of the content of first principles—and not to every application of practical reason. And, as I have already explained, this much is perfectly compatible with the Conventional Interpretation and Rawls's own appeal to such a view from within moral theory.

One could argue that the view in *KCMT* would be more coherent if it characterized every application of practical reason as it does the determination of the content of first principles—i.e., without appealing to any special ontology (stance-dependent or otherwise). Whether this is the case, however, depends on what it means to be external to the activity of practical reason itself and whether the stance-dependent truth-makers generated by the construction procedure count as external in this sense. Settling these questions would take us beyond the scope of the current essay. More importantly, this would also appear to be unnecessary.

Despite the popularity of such a view amongst contemporary defenders of Kantian constructivism, there is good textual evidence for thinking that Rawls does not share it. It is again worth noting that the dissertation is an underappreciated resource in this respect. I will conclude by briefly presenting Rawls's worries about metaphysics, as stated in the dissertation, and arguing that they in no way undermine the Conventional Interpretation

---

[48]Although, to my knowledge, this characterization of constructivism has not been offered as an interpretation of the lectures, it has been defended independently in Engstrom (2013) and Bagnoli (2013a). While Korsgaard (2003) explicitly rejects a knowledge-based conception of constructivism, familiar readers will recognize obvious parallels between the view developed there and the second interpretation I have in mind here.

or the plausibility of a correspondence theory of truth for morality.

Rawls is explicit in the dissertation that his apprehension concerns metaphysics in a relatively narrow sense.[49] In particular, he thinks it is a mistake to justify ethical judgments by 'appeal to exalted entities' (Rawls 1950, 318–20). Although he lists a number of different candidates for such entities from the historical record (including divine commands, the General Will, *Weltgeist*, Humean moral sense, Kantian transcendental psychology, perfectionist human nature: see Rawls 1950, 320–27), he would appear most concerned to respond to the kind of non-naturalist ethical realism one finds in the works of G. E. Moore, W. D. Ross, and C. D. Broad (Rawls 1950, 279–83, 327n).

The problem with all such views, according to Rawls, is that the justificatory arguments they present fail to appreciate the special authority of ethical judgments. Each, in their own way, involves deferring to an external source of authority, 'such and such an entity, in its nature or behavior, sanctions so and so, therefore, so and so ought to be' (Rawls 1950, 319). But Rawls objects that this would violate Hume's Law—since, on all such views, an 'ought' is inferred from an 'is'. The 'deeper' problem with this, as Rawls explains, is that such a violation conflicts with the conception of the person as free and autonomous.

> The actual separation of the 'is' and the 'ought,' though embodied in the logic of everyday language, actually has this deeper root: namely, to keep the collective sense of right free from being necessitated to adopt any criteria of right and wrong except those to which it freely consents. Nothing that is, simply because it is, ought to be, no matter what how [sic] exalted it may be; what ought to be is only decided by what the collective sense of right freely

and spontaneously believes ought to be as this fact is shown by its conformity to justifiable principles. (Rawls 1950, 319–30)

It is clear from this and other nearby passages that Rawls's worry only concerns an appeal to certain kinds of metaphysical views (see also Rawls 1950, 330), ones according to which the source of authority is external to 'what the collective sense of right freely and spontaneously believes.' Rawls explicitly makes exception for views that justify ethical judgments by appeal to facts about what would be accepted by someone possessing and exercising the relevant moral personality.

> The view here may be stated as follows: ethical principles are 'metaphysikfrei' with respect to all of what is, or may be, *except that portion of ascertainable fact which is expressed by saying that they are accepted by reasonable men after criticism and reflection, and in the light of their common sense judgments . . .* Thus the dichotomy between what is, and what ought to be, is here maintained. The interpretation put upon it is simple but important: it registers our determination not to be intimidated by power; it records our effort to make the collective sense of right the final test of right and wrong. (Rawls 1950, 334–35, emphasis added)

Rawls does not reject appeal to all facts or truth-makers, only the 'exalted' ones.[50] Whereas any appeal to facts will involve some form of ontological commitment, Rawls is only worried about the positing of a special class of entities, along the lines of a non-naturalist ethical realism or a divine command theory. By contrast, it is not obvious that the kinds of facts, or truth-makers, that Rawls appeals to here in the dissertation, or later in the lectures, requires any such special ontology. Rather, all that is required are facts about moral psychology and persons who regard themselves in certain ways. This much would appear to

---

[49]Weithman would appear to agree with this claim, though he also thinks 'it would miss the point Rawls is trying to make' (2010, 34). Again, I do not think that Rawls intended to focus on these metaphysical aspects of Kantian constructivism, but I do think that they are essential for understanding the view he presents in the lectures.

[50]Earlier Rawls explains that '*metaphysikfrei*' does not mean independent of all facts, but 'valid irrespective of the truth or falsity of a particular body of empirical and metaphysical propositions which, in each case, to be precise, should be specified' (Rawls 1950, 135–36).

be fully accounted for by a naturalistic ontology. This is just to say that Rawls's worries about metaphysics and authority are perfectly compatible with an appeal to the Conventional Interpretation of Kantian constructivism. They in no way require the kind of metaphysical quietism Rawls presents in *PL*, nor even the more thoroughgoing, and ontologically non-committal, constructivism about practical reason that is preferred by some contemporary Kantians. In other words, the available textual evidence in no way rules out or undermines a characterization of constructivist moral personality that involves appeal to a correspondence theory of truth.

Moreover, there is even some reason for thinking that such a characterization would better serve Rawls's aim of contrasting the constructivist self-conception with that of rational intuitionism. Recall that Rawls distinguishes constructivism from rational intuitionism by appealing to an explanatory notion of priority—one that concerns the kinds of facts that *explain* why moral statements are true or false. This priority relation would appear best accounted for from within the framework of a correspondence theory of truth. Although we may assume that any plausible view of truth must be able to accommodate the stance-dependence/independence distinction at the level of properties or concepts, it is doubtful that any non-correspondence theory would support assigning this distinction a substantive role in explaining *the truth* of moral principles.[51] For example, if Rawls

---

[51]But why think that the relevant explanatory contrast concerns truth and not the properties themselves? For example, a commitment to minimalism would still allow one to distinguish views according to which the instantiation of a moral property, like wrongness, is explained by its dependence on stances from those according to which it is explained by stance-independent facts. Rawls could have presented the contrast this way, but this is not the contrast we get in *KCMT* or any of the other texts in which Rawls discusses rational intuitionism. As I have already established, Rawls's takeaway from Kant is that rational intuitionism gives the wrong account of *why moral statements are true or false* (again, compare 1989, 95–96.); the criticism is not that it unnecessarily or mistakenly relies on a substantive notion of truth. Hence, the contrast we

assumed a minimal view of truth—one that did not involve any substantive explanation of what it is for a statement to be true, then it would be unclear how Rawls could coherently appeal to this kind of priority. On such an account, the contrast between rational intuitionism and constructivism would appear lost.[52] In response, one may retreat to the position that the lectures are simply underdetermined on this point and that the best interpretation is a kind of agnosticism about truth. Such an interpretation is certainly compatible with the textual evidence. Nevertheless, this view is neither forced upon us by the text nor does it underscore the contrast between the constructivist's and intuitionist's forms of self-regard to the same extent that the correspondence theory arguably would.

6. Conclusion: *KCMT* Resituated

I have laid out the best possible case for reading the constructivism of *KCMT* as appealing to a stance-dependent, cognitivist success theory at the level of moral theory. I have presented evidence that speaks for this claim, as well as evidence that cuts against it. The two biggest challenges to my interpretation are, on the one hand, Rawls's apparent disavowal of moral truth and, on the other hand, his apparent disavowal of metaethics. In response to the first challenge, I have argued that Rawls is best understood as introducing 'reasonable' as a way of flagging a difference in the nature of the truth-makers for moral statements.

---

get in *KCMT* is best read as concerning moral *truth* as explanandum, not moral properties or concepts.

[52]Note, again, that Rawls also explicitly rules out an epistemic conception of truth for morality (*KCMT*, 565–56). Rawls's reason for voicing this rejection is that some commentators have interpreted Kant's moral theory as requiring such an alternative. Rawls both rules out that the nature of moral truth differs from theoretical truth and flags a potential point of divergence between his constructivism and Kant's—should such an interpretation of Kant prove correct.

On my reconstruction, Rawls does not reject the claim that moral statements are truth-apt or, alternatively, accept the claim that truth has a different nature in moral discourse. In response to the second objection, I have argued that, despite worries about exalted entities and countervailing evidence in later work, the constructivist self-conception Rawls presents in *KCMT* is best read in terms of the Conventional Interpretation and not as a species of metaethical quietism or, alternatively, a form of cognitivism that grounds practical knowledge without ontological commitment.

The interpretation I have defended is conventional in the sense that it attributes to *KCMT* a common way of understanding the metaethical features appealed to in Kantian constructivism. But there are also two important senses in which the reading of the lectures I have presented is anything but conventional.

First, on my interpretation, there is a way in which the Conventional Interpretation gets things wrong. Although Kantian constructivism involves the metaethical features one finds in the Conventional Interpretation, it is neither reducible to these features nor does it theoretically commit one to these views outside of the perspective of moral theory. It is not reducible to these features because moral theory is not primarily concerned with metaethics; rather, these features only enter into a characterization of constructivism indirectly, as part of the description of the constructivist self-conception and the form of self-regard they contribute to. It does not theoretically commit one to the Conventional Interpretation because, regardless of one's views in moral theory, it is possible to hold any number of positions in metaethics. Again, however, this is not to say that one's views in moral theory have *no* bearing on one's views in moral philosophy, more broadly. Indeed, Rawls is keen to stress the mutual dependence of our beliefs at all levels of generality, concerning all parts of philosophy.

Second, the reading I have defended requires that we radically resituate the role of the lectures in the development of Rawls's thinking. To the extent that philosophers have ventured to comment on the role of the lectures in Rawls's broader corpus, they almost uniformly describe them as a 'transitional stage' between *TJ* and *PL*.[53] And, of course, it is easy to see why. The lectures were delivered in 1980, nine years after the first publication of *TJ* and thirteen years prior to *PL*. There is also considerable overlap across these three texts in Rawls's presentation of justice as fairness. Those who argue that the lectures draw out and make explicit certain inconsistencies in Rawls's argument for the stability of justice as fairness make a very convincing case (Freeman 1999, 2007; Weithman 2010). Nevertheless, if what I have argued is correct, the real significance of the lectures is better appreciated in relation to an altogether different set of texts. *KCMT* marks the culmination of Rawls's response to Sidgwick and the development of his views on moral theory, work that started in his doctoral dissertation and that received further expression and refinement in 'Outlines of a Decision Procedure' and 'The Independence of Moral Theory'. This project surely extends beyond these works, informing Rawls's contribution to the history of philosophy, as well as *TJ* and *PL*. But its core is to be found in these four texts. It is finally in *KCMT* that one can fully appreciate Rawls's critique of *Methods* and see the structure of and possibilities for an approach to moral theorizing that assigns primacy to conceptions of moral personality.

[53]Freeman (1999, ix) is most explicit about this; see also his (2007, 5–6). Although Weithman (2010) is not much concerned with *constructivism per se* (the term 'constructivism' only occurs twice: see 10, 266), there are ample references to the lectures which characterize the work as a stage of Rawls's transition from *TJ* to *PL* (see, e.g., 244–45, 247, 265–66, 270, 281–82, 284–85, 290–91). The one exception, as far as I am aware, can be found in Barry (1989). Barry is not explicit on this point, in part because his real focus is on *TJ*, but he does suggest that he would be sympathetic to the line I take here when he writes: 'As a matter of intellectual biography, *A Theory of Justice* may be seen as a point of transition between the coherentism of Rawls's early article (based on his Ph.D. dissertation) "Outline of a Decision Procedure for Ethics," dating from twenty years before the book, and the constructivism of the Dewey Lectures and subsequent articles' (278–79).

Acknowledgements

Nathaniel Jezzi
Lafayette College
jezzin@lafayette.edu

References

Bagnoli, Carla, 2013a. 'Constructivism about Practical Knowledge.' In Bagnoli (2013b), pp. 153–82.

———, ed., 2013b. *Constructivism in Ethics*. Cambridge: Cambridge University Press.

Barry, Brian, 1989. *A Treatise on Social Justice*, vol. I, *Theories of Justice*. London: Harvester-Wheatsheaf.

Blackburn, Simon, 1998. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.

Brink, David, 1987. 'Rawlsian Constructivism in Moral Theory.' *Canadian Journal of Philosophy* 17: 71–90.

———, 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Cohen, Joshua, 2009. 'Truth and Public Reason.' *Philosophy & Public Affairs* 37: 2–42.

Darwall, Stephen, Allan Gibbard and Peter Railton, 1992. 'Toward a Fin de siecle Ethics: Some Trends.' *Philosophical Review* 101: 115–89.

Dorsey, Dale, 2012. 'A Puzzle for Constructivism and How to Solve It.' In Lenman and Shemmer (2012), pp. 99–118.

Engstrom, Stephen, 2013. 'Constructivism and Practical Knowledge.' In Bagnoli (2013b), pp. 133–52.

Enoch, David, 2009. 'Can There Be a Global, Interesting, Coherent Constructivism about Practical Reason.' *Philosophical Explorations* 12: 319–39.

Fitzpatrick, William, 2005. 'The Practical Turn in Ethical Theory: Korgaard's Constructivism, Realism and the Nature of Normativity.' *Ethics* 115: 651–91.

Freeman, Samuel, ed., 1999. *John Rawls: Collected Papers*. Cambridge, MA: Harvard University Press.

———, ed., 2003. *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.

———, 2007. *Justice and the Social Contract: Essays on Rawlsian Political Philosophy*. Oxford: Oxford University Press.

Galvin, Richard, 2010. 'Rounding Up the Usual Suspects: Varieties of Kantian Constructivism in Ethics.' *Philosophical Quarterly* 61: 16–36.

Gibbard, Allan, 1999. 'Morality as Consistency in Living.' *Ethics* 110: 140–64.

———, 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Hill, Thomas, 1989. 'Kantian Constructivism in Ethics.' *Ethics* 99: 752–70.

———, 2011. 'Kantian Constructivism as Normative Ethics.' *Oxford Studies in Normative Ethics* 1: 26–50.

Hussain, Nadeem, 2012. 'A Problem for Ambitious Metanormative Constructivism.' In Lenman and Shemmer (2012), pp. 180–94.

Hussain, Nadeem and Nishi Shah, 2006. 'Misunderstanding Metaethics: Korsgaard's Rejection of Realism.' *Oxford Studies in Metaethics* 1: 265–94.

Irwin, Terence, 2009. *The Development of Ethics: A Historical and Critical Study*, vol. III. Oxford: Oxford University Press.

James, Aaron, 2007. 'Constructivism about Practical Reasons.' *Philosophy and Phenomenological Research* 74: 302–25.

———, 2012. 'Constructing Protagorean Objectivity.' In Lenman and Shemmer (2012), pp. 60–80.

Korsgaard, Christine, 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

———, 2003. 'Realism and Constructivism in Twentieth-Century Moral Philosophy.' *Journal of Philosophical Research* APA Centennial Supplement: 99–122.

LeBar, Mark, 2008. 'Aristotelian Constructivism.' *Social Philosophy and Policy* 25: 182–213.

Lenman, James, 2010. 'Humean Constructivism in Moral Theory.' *Oxford Studies in Metaethics* 5: 175–93.

———, 2012. 'Expressivism and Constructivism.' In Lenman and Shemmer (2012), pp. 213–25.

Lenman, James and Yonatan Shemmer, eds., 2012. *Constructivism in Practical Philosophy*. Oxford: Oxford University Press.

Lynch, Michael, 2009. *Truth as One and Many*. Oxford: Clarendon Press.

Mackie, J. L., 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin Books.

Milo, Ronald, 1995. 'Contractarian Constructivism.' *Journal of Philosophy* 92: 181–204.

Moore, G. E., 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

O'Neill, Onora, 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press.

———, 2003a. 'Constructivism in Rawls and Kant.' In Freeman (2003), pp. 347–67.

———, 2003b. 'Constructivism vs. Contractualism.' *Ratio* n.s. XVI 4: 319–31.

Rawls, John, 1950. A Study in the Grounds of Ethical Knowledge: Considered with Reference to Judgments on the Moral Worth of Character. Ph.D. dissertation, Princeton University.

———, 1951. 'Outline of a Decision Procedure for Ethics.' *The Philosophical Review* 60: 177–97.

———, 1971/1999. *A Theory of Justice*, 2nd ed.. Cambridge, MA: Harvard University Press, 1999. (First edition 1971.)

———, 1974–75. 'The Independence of Moral Theory.' *Proceedings and Addresses of the American Philosophical Association* 48: 5–22.

———, 1980. 'Kantian Constructivism in Moral Theory.' *Journal of Philosophy* 77: 515–72.

———, 1985. 'Justice as Fairness: Political not Metaphysical.' *Philosophy & Public Affairs* 14: 223–51.

———, 1989. 'Themes in Kant's Moral Philosophy.' In *Kant's Transcendental Deductions*, edited by Eckhard Forster, pp. 81–113. Stanford: Stanford University Press.

———, 1993/1996. *Political Liberalism*, 2nd ed. New York: Columbia University Press, 1996. (First edition 1993.)

———, 2000. *Lectures on the History of Moral Philosophy*. Cambridge, MA: Harvard University Press.

Ridge, Michael, 2012. 'Kantian Constructivism: Something Old, Something New.' In Lenman and Shemmer (2012), pp. 138–58.

Scanlon, T. M., 1982. 'Contractualism and Utilitarianism.' In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams, pp. 103–28. Cambridge: Cambridge University Press.

Schafer, Karl, 2015a. 'Realism and Constructivism in Kantian Metaethics (1): Realism and Constructivism in a Kantian Context.' *Philosophy Compass* 10: 690–701.

———, 2015b. 'Realism and Constructivism in Kantian Metaethics (2): The Kantian Conception of Rationality and Rationalist Constructivism.' *Philosophy Compass* 10: 702–13.

Shafer-Landau, Russ, 2003. *Moral Realism: a Defence*. Oxford: Oxford University Press.

Sidgwick, Henry, 1907/1966. *The Methods of Ethics*, 7th edition. New York: Dover Publications, 1966. (Edition first published 1907.)

Silk, Alex, 2015. 'Nietzschean Constructivism: Ethics and Metaethics for None and All.' *Inquiry* 58: 244–80.

Street, Sharon, 2008. 'Constructivism about Reasons.' *Oxford Studies in Metaethics* 3: 207–45.

———, 2010. 'What is Constructivism in Ethics and Metaethics?' *Philosophy Compass* 5: 363–84.

Timmons, Mark, 1999. *Morality without Foundations: A Defense of Ethical Contextualism*. Oxford: Oxford University Press.

———, 2003. 'The Limits of Moral Constructivism.' *Ratio* 16: 391–423.

Weithman, Paul, 2010. *Why Political Liberalism? On John Rawls's Political Turn*. Oxford: Oxford University Press.