Rawls and Kantian Constructivism

ALEXANDER KAUFMAN

School of Public and International Affairs, University of Georgia

Abstract

John Rawls's account of Kantian constructivism is perhaps his most
striking contribution to ethics. In this paper, I examine the relation
between Rawls's constructivism and its foundation in Kantian intuitions.
In particular, I focus on the progressive influence on Rawls's approach of
the Kantian intuition that the substance of morality is best understood as
constructed by free and equal people under fair conditions. Rawls's focus
on this Kantian intuition, I argue, motivates the focus on social contract
that grounds both his accounts of the original position and of reflective
equilibrium. Critics, including Onora O'Neill and Larry Krasnoff, object
that Rawls's view distorts various aspects of Kantian moral reasoning.
I argue that these objections (i) exaggerate the distinctions between
Kant's and Rawls's decision procedures and (ii) reflect an unnecessarily
constricted view of Kant's moral thought.

John Rawls's account of Kantian constructivism is perhaps his most
striking contribution to ethics. The notion of constructivism developed
in this account has, in fact, been described as a new possibility in ethical
and political theory (see Krasnoff 1999: 85) because it employs anti-
realist resources to address concerns raised by moral sceptics and
pluralist critics of liberalism. A constructivist approach holds that
moral propositions are right or justified when they are consistent with
acceptable moral principles, and moral principles are acceptable when
they are the product of an appropriately designed decision procedure.
A Kantian form of moral constructivism, in Rawls's account, is dis-
tinguished by the central role that such a theory assigns to a decision
procedure that constitutes a procedural interpretation of Kant's ideas
regarding moral reasoning and autonomy. The role assigned to such a
procedure, Rawls argues, reflects Kant's view that the substance of
morality is fixed neither by an independently existing order of values

nor by special features of human psychology. Rather, that substance is best understood as constructed by free and equal people under fair conditions. This constructivist interpretation of Kant's moral thought has had a decisive influence on Kant scholarship in the Anglo-American tradition. Although many scholars in the continental tradition continue to view Kant as a rational intuitionist, even 'a covert moral realist' (O'Neill 2008: 54), Kant is firmly established among Anglo-American commentators as an anti-realist and even a formalist moral thinker.

The constructivist representation of moral judgements formed in the context of a concretely described decision procedure and informed by a general knowledge of facts about human society has led to the perception among many commentators that Rawls's constructivism is grounded in specific factual information or in principles or values that are not principles or values of justice (see Cohen 2008: 278–86). Thus, G. A. Cohen claims Rawls's constructivism is 'fact-infested' (Cohen 2008: 287) and, in addition, grounds its judgements in strategic and other considerations that are irrelevant to justice (Cohen 2008: 280), while Aaron James asserts that Rawls assumes that constructivist moral judgement is authoritative only when 'grounded in independent judgments about what kind of social practices exist and what kinds of agents participate in them' (James 2005: 282).

Both of these views, however, misrepresent the character of Rawls's account. Cohen and James, like many other commentators on Rawls's work, seem to assume that a moral judgement in Rawls's account 'either articulates a description of some fact or is a disguised version of some alternative use of language' (Korsgaard 2008: 309). Thus, the confusion in both arguments derives from a failure to take seriously the centrality in Rawls's account of the Kantian intuition that moral judgements are grounded neither in facts nor in independent and pre-existing norms or principles, but rather in a process of reasoning.

In examining this strand of Rawls's work, it is important to distinguish Kantian constructivism from the substantive accounts of justice that Rawls develops in *A Theory of Justice* and *Political Liberalism*. While Rawls's constructivism works from many of the same assumptions that ground his substantive political theory, Kantian constructivism provides an account of the structure of moral reasoning that is independent of both justice as fairness and political liberalism. Rather than providing or supplementing an account of a substantive moral or political conception, Kantian constructivism develops an approach to assessing the

reasonableness of moral judgements that can be employed *to evaluate* substantive moral conceptions.

It is equally important to distinguish Kantian constructivism from Rawls's account of *political* constructivism. Although political constructivism constitutes the view that is most relevant to the final version of Rawls's theory of political justice, Kantian constructivism represents a distinct strand of Rawls's thought with significance independent of its relation to Rawls's theories of justice as fairness and political liberalism. In particular, Kantian constructivism (i) develops a distinctly Kantian approach to justification in ethics; (ii) applies to a wide range of moral questions rather than being limited to issues relating political and social institutions; and (iii) has achieved a broad influence among contemporary moral philosophers (see James 2007; Hill 2002; Scanlon 1998; Korsgaard 1996).

In this paper, I will examine the relation between Rawls's constructivism and its foundation in Kantian intuitions. In particular, I will focus on the progressive influence on Rawls's approach of the Kantian intuition that the substance of morality is best understood as constructed by free and equal people under fair conditions and is not determined by appeals to the authority of a pre-existing and independent order of values. Rawls's focus on this Kantian intuition, I will argue, motivates the focus on social contract that grounds his accounts both of the original position and of reflective equilibrium. I will first briefly sketch Rawls's general approach to moral justification and the status of constructivism within that approach. In section 2, I will examine the progressive influence of Kantian intuitions on Rawls's political thought and account of constructivism, both in his development of ideas from social contract doctrine and in his extension of ideas from 'Outline of a Decision Procedure for Ethics'. In section 3, I will discuss and reject arguments in the current constructivist literature that locate the authority of Rawls's constructivism in facts or in non-moral values or elements of existing practices. In section 4, I will examine leading objections to Rawls's constructivist interpretation of Kant. Finally, I will conclude by discussing (i) the status of Kantian constructivism within Rawls's approach to moral justification and (ii) its relation to political liberalism.

1. Moral Justification

While constructivism constitutes a salient component of Rawls's account of moral justification, it nevertheless represents a subordinate strand of that account. Rawls's most general standard of moral and

political justification requires that a moral proposition is justified if and only if it matches the considered judgements of a reasonable and rational person on due reflection in reflective equilibrium. In due reflection, the person reasons from premises that are (i) 'widely shared but weak' (Rawls [1971] 1999: 16) and (ii) judged to be reliable when viewed from 'conditions favorable for deliberation and judgment in general' (Rawls [1971] 1999: 42). Premises that satisfy this standard constitute *considered judgements* 'in which our own moral capacities are most likely to be displayed without distortion' (Rawls [1971] 1999: 42). During due reflection, the person 'attempts to organize the basic ideas and principles implicit' in his or her considered judgements into a coherent moral or political view (Rawls [1993] 1996: 8) and aims to reach a state of reflective equilibrium in which his or her 'general convictions, first principles, and particular judgements are … in line' (Rawls [1993] 1996: 384, n. 16).

Rawls's account of due reflection in fact extends his earlier efforts – efforts that reflect a number of non-Kantian influences – to describe criteria for evaluating the validity of proposed moral and political judgements. According to the arguments developed in 'Outline of a Decision Theory for Ethics', a moral judgement is valid if and only if it would be accepted by competent moral judges, and a principle is valid if and only if it shows 'a capacity to hold its own' (Rawls 1951: 11) against a subclass of relevant considered judgements. This account of the validity of moral judgements, in its attempt to apply 'the relevant requirements of practical reason' (Rawls [1993] 1996: 90) to determine which moral judgements best satisfy the criteria of a rational method of ethics, reflects the influence of Henry Sidgwick's *The Methods of Ethics*. Rawls's assumption, in 'Two Concepts of Rules', that reflection about questions of justice must identify principles that apply primarily to basic institutions reflects the influence of David Hume's *Treatise of Human Nature* and of Ludwig Wittgenstein's idea of a social practice.[1]

Rawls's reflections regarding the nature of valid moral and political judgements lead him to conclude that such judgements are properly viewed as conclusions derived through the operations of an acceptable deliberative procedure. In developing this view, Rawls is increasingly drawn to the intuition – which he attributes to Kant – that moral reasoning is 'part of the general theory of rational choice'[2] applied to the problem of securing reasonable and mutually justifiable social relations. Once Rawls has reached this conclusion, the idea of social contract – in particular, Kant's account of social contract – plays an increasingly central role in Rawls's account of justification in ethics.

2. The Kantian Influence

While the influence of Kant is not evident in Rawls's earliest published work (see Rawls 1951, 1955), Rawls begins to refer to the central importance of Kant's thought in 'Justice as Fairness', and discussions of the central significance of Kant's thought are increasingly salient in Rawls's work between 1958 and 1980. Rawls is careful, however, to emphasize that his work does not attempt to offer exegesis of Kant's views. Rather, it develops Kantian themes in a manner that makes Rawls's work 'closer to [Kant's] view than to the other traditional moral conceptions that are appropriate for use as benchmarks of comparison' (Rawls 1980: 305).

Rawls's working out of Kantian themes increasingly affects both his account of justification in ethics and his development of his substantive account of justice during the period between the publication of 'Justice as Fairness' and 'Kantian Constructivism in Moral Theory'. Most fundamentally, Rawls is influenced by Kant's view that a moral action derives its worth from the principle of volition that motivates the action rather than from any particular end attained.[3] In the first part of the *Groundwork*, Kant argues that 'an action from duty has its moral worth *not in the purpose* to be attained by it ... but merely [in] the *principle of volition* in accordance with which the action is done' (Kant 1996: 55; *G* 4: 400).[4] In this and similar passages, Kant argues that moral reasoning is characterized not by the pursuit of certain ends or goals, but by the reason for action contained in the underlying principle of volition. The analysis of moral judgement, for Kant, becomes an analysis of non-instrumental principles of volition and arguments constructed from those principles.

If the nature of the reason grounding an act determines its moral quality, then the moral quality of an act can be assessed by determining whether reasonable persons under suitable conditions would have sufficient reasons to consent to the act. Kantian intuitions thus lead Rawls to focus his moral analysis around the notion of rational consent as developed in Kant's theory of social contract. Social contract, Rawls notes in 'Justice as Fairness', 'does express, suitably interpreted, an essential part of the concept of justice' (Rawls 1958: 71). An examination of the evolution of Kant's progressive influence on Rawls in fact suggests that Rawls's constructivism evolves directly from his interpretation of Kant's theory of social contract and not from an interpretation of the categorical imperative procedure. In addition, Rawls's extension of the Kantian social contract argument provides a basis for the extension of the

non-Kantian strand of Rawls's account of justification in ethics developed in 'Outline of a Decision Procedure for Ethics'. Finally, the notion of a social contract doctrine as a form of rational choice theory suggests to Rawls the idea of employing a hypothetical social contract model as a procedure that specifies the relation between a particular conception of the person and a conception of justice.

Constructivism and Social Contract

Rawls first notes the significance of Kant's theory of social contract for his approach in 'Justice as Fairness', where he suggests that 'Kant was not far from wrong when he interpreted the original contract merely as an idea of reason' (Rawls 1958: 71). While legal and political obligation do not literally originate in any form of original agreement, social contract constitutes 'an ethical idea applicable to social arrangements irrespective of the question of origins' (Rawls 1971: 223). Just as ideas of reason generally secure the greatest unity and extension for the concepts of the understanding, the idea of social contract can be employed 'to clarify the concept of justice' by representing its general unifying ground (Rawls 1958: 59). This view of justice as grounded in the notion of social contract, in fact, follows naturally from Rawls's view that the question of justice 'arises once the concept of morality is imposed upon mutually self-interested agents, similarly circumstanced' (Rawls 1958: 59). If justice is a conception that self-interested agents create to regulate their joint interactions, then the metaphor of a contract seems appropriate to represent the kind of jointly acceptable standard for behaviour that is required.

In presenting the idea of social contract as a general unifying ground for the concept of justice, Rawls follows Kant's view that the idea of an 'original contract' – by which a people 'unites into a society' by 'establishing a *civil constitution*' (Kant 1996: 290; TP 8: 289) through the exercise of the 'general (united) will of the people' (Kant 1996: 295; TP 8: 295) – constitutes 'the touchstone of any public law's conformity with right' (Kant 1996: 297; TP 8: 297). Rawls also follows Kant in viewing the idea of social contract as uniquely well-suited for this role *because* it expresses the fundamental idea that is essential to judgments of justice – the idea of fairness (Rawls 1958: 47). As Kant notes, public laws generated by the general will of the people must be 'incapable of doing wrong to anyone', since 'all decide about all, hence each about himself' (Kant 1996: 295; TP 8: 295). It is this aspect of the concept of justice. Rawls argues, that is neglected by utilitarian theory – utilitarianism improperly applies 'the principle of choice for one man' to society

(Rawls 1967: 132). Social contract theory, Rawls argues, therefore comes closer than utilitarian theory to expressing an essential part of the concept of justice (Rawls 1958: 71).

Once principles of justice are viewed as the product of joint agreement, the construction of a conception of justice requires precise accounts of (i) fair conditions for agreement and (ii) genuine, free, and informed consent. Rawls presents definitive accounts of both of these elements of his theory in *A Theory of Justice*, and commentators have generally assumed that Rawls models the decision procedure that incorporates these elements – the original position – directly upon the categorical imperative procedure. An examination of the development of the Kantian theme of social contract in Rawls's early articles, however, contradicts this standard view. Rawls develops his constructivist approach to justification by working directly from Kant's theory of social contract and only later takes note of the relation between his account and the categorical imperative procedure.

Rawls first offers an account of a contractarian approach to justification in 'Justice as Fairness'. In this paper, Rawls argues that the choice of acceptable principles of justice may be viewed as the product of willing agreement among rational agents under conditions that meet 'the standards which parties could accept as fair' (Rawls 1958: 63). Rawls develops the idea by discussing a hypothetical account of deliberation about justice among rational agents whose allegiance to their existing practices is grounded in the role of those practices in securing their rational advantage (Rawls 1958: 56–60). In the process of deliberation, participants propose principles of justice to regulate current disputes in which they are engaged. A proposed principle becomes binding only of it is acceptable to all participants, and an accepted principle binds all members of society in all relevantly similar future cases. While the participants are thus fully informed regarding their situated interests, Rawls designs the choice situation to prevent cases in which some parties may be 'taken advantage of, or forced to give in to, claims which they do not view as legitimate' (Rawls 1958: 59) by requiring that: (i) all participants must consent to a principle before it can be accepted; and (ii) all accepted principles must apply to all persons in precisely the same way.

As in Rawls's later account of the original position, the circumstances of choice are designed to be fair and, in particular, to eliminate bargaining advantage from the factors affecting the choice of principles. It is

important, however, to note that Rawls's initial employment of the social contract notion does not lead him to suggest the idea of a hypothetical decision procedure similar to the categorical imperative procedure. Rather, Rawls develops the idea of social contract through an account of quasi-historical deliberation among members of a society 'among whom a certain system of practices is *already* well established' (Rawls 1958: 52). Unlike choosers in the original position, participants retain full knowledge of their situated interests, such as social position and wealth. While the process is well designed to motivate choosers to avoid the choice of principles that *explicitly* discriminate against persons or groups, the process is not structured in a way that guards against the possible adoption of rules or principles with concealed differential impacts. Thus, Rawls's sketch of deliberations about justice in 'Justice as Fairness' presents the notion of social contract as a regulative idea to guide the judgement of situated participants in the legislative process, but does not develop the idea as the basis for (i) a decision process that embodies the value of impartiality or (ii) a hypothetical decision procedure.

In papers published between 1958 and 1967, Rawls relies upon this contractarian account of rational choice under fair circumstances as his model of just deliberation about principles of justice (see Rawls 1963a, 1963b, 1964: 123). In 'Distributive Justice', however, Rawls significantly extends his account of the theory of social contract, which he continues to describe as the 'natural alternative to utilitarianism'.[5] In particular, Rawls develops a precise account of the fair choice position under which acceptable principles of justice may be jointly chosen by free rational persons. This position, Rawls now states, is an 'original position of equality' in which 'no one knows his position in society, nor even his place in the distribution of natural talents and abilities' (Rawls 1967: 131–2). In this choice position, '[a] veil of ignorance prevents anyone from being advantaged or disadvantaged by the contingencies of social class and fortune' (ibid. 132). In 'Distributive Justice', then, Rawls develops the procedural implications of his understanding of social contract doctrine. While Rawls's account of a fair choice position in this paper corresponds closely to his definitive view of the original position in *A Theory of Justice*, however, 'Distributive Justice' still provides an account of the conditions under which the choice of principles of justice may be viewed as the product of a quasi-historical agreement, and not as the product of a purely hypothetical decision procedure.

Finally, in *A Theory of Justice*, Rawls further refines his account by eliminating the notion of a quasi-historical agreement. Rawls retains

from 'Distributive Justice' his account of the informational constraint imposed by the veil of ignorance, but describes the account of choice subject to that constraint as entirely hypothetical. Thus, Rawls's account of the original position emerges as an interpretation of Kant's idea of *social contract* and does not constitute an attempt to apply the categorical imperative to social problems.

Social Contract and Reflective Equilibrium

While the account of a fair choice position in 'Distributive Justice' develops the Kantian contractarian strand in Rawls's argument, this account of a choice position grounded in an interpretation of the idea of social contract also provides a basis for the extension of the non-Kantian strand of Rawls's account of justification in ethics developed in 'Outline of a Decision Procedure for Ethics'. As noted above, that paper argues for a specific criterion for assessing the validity of moral judgements that is clearly related to Rawls's account of reflective equilibrium in *A Theory of Justice*. In both accounts, considered judgements constitute the raw material of valid moral judgements, and a valid judgement employs this raw material in a carefully specified process that is designed to eliminate – as far as possible – the usual sources of error and distortion in moral reasoning.

One essential element of Rawls's later approach is, however, not present in the earlier account of moral reflection. In the later approach, moral agents generate an account of the 'significant bounds' that their considered judgements of justice, taken together, 'impose … on acceptable principles of justice' (Rawls [1971] 1999: 16), and these bounds are represented in the features of the original position. Once an acceptable account of the original position is specified, the parties (i) select acceptable principles of justice from that standpoint; (ii) compare the judgements required by these principles in particular cases to those required by their more specific considered judgements; and (iii) refine both their principles and their judgements until these produce the same results when applied. Rawls describes the point at which their principles and judgements coincide as *reflective equilibrium*.

The process of achieving reflective equilibrium imposes two conditions that refine both the set of considered judgements and the principles: (i) the set of considered judgements must be consistent among themselves; and (ii) the set of considered judgements must be consistent with principles that 'extend them in an acceptable way' (Rawls [1971] 1999: 17). The second of these conditions imposes a significant constraint on

the set of acceptable principles – such principles must specify grounds justifying the original set of considered judgements that are sufficiently general to support moral judgements that *extend* that original set of considered judgements. This condition is not imposed by the decision procedure of 'Outline of a Decision Procedure for Ethics'. The process of achieving reflective equilibrium thus provides Rawls's later account of justification in ethics with significant critical leverage that is lacking in his earlier account.

It is the incorporation of the account of a choice position grounded in an interpretation of the idea of social contract, along with his account of the process through which the choice position is derived and justified, that provides Rawls's later approach to justification in ethics with this critical leverage. Rawls's interpretation of Kant's theory of social contract thus provides Rawls with one of the central features of his account of reflective equilibrium.

Social Contract and the Kantian Conception of the Person

The distinguishing feature of the fair choice position employed in Kantian constructivism, Rawls states, is its specification of 'a particular conception of the person as an element of [the] reasonable procedure' that determines the choice of principles (Rawls 1980: 304). Thus, it is not the structure of the procedure itself, but rather the relation between that structure and a Kantian conception of the person, that is most fundamental to Kantian constructivism. While Rawls does not set out an explicit textual basis for this characterization of the Kantian conception of the person, Kant's political and moral theory provides support for Rawls's characterization.

Rawls's account of the Kantian conception of the person takes as its foundation the fundamental qualities attributed to the person as a citizen in Kant's political philosophy. In the 'Theory and Practice' essay, Kant asserts that a state that is in conformity with right must be founded on a view of persons as free, equal and independent (TP 8: 290–1, see R 237). Persons are viewed as *free* in the sense that each may 'seek his happiness in the way that seems good to him' (TP 8: 291), as *equal* in their authority to exercise coercive rights against one another, and as *independent* in their authority as co-legislators (Kant 1996: 291–5; TP 8: 291–4). Rawls's view of the senses of *freedom* and *equality* that must be embodied in an acceptable decision procedure follows Kant fairly closely, while Rawls incorporates *independence* into his notion of freedom. Freedom, in Rawls's account, refers to the status

of members of society as 'self-originating sources of claims' (Rawls 1980: 334); equality refers to the fact that each member 'has the same rights and powers'; and 'freedom as independence' refers to the grounds that persons have to consent to legislation under the informational constraints of the original position (Rawls 1980: 335).

The third and fourth qualities that Rawls incorporates in the Kantian conception of the person – the abilities to act reasonably and rationally – Rawls draws from the *Groundwork*. Rawls employs the terms *reasonable* and *rational* to characterize the two forms of practical reason, pure and empirical. An employment of practical reason is *reasonable* if it reflects 'willing[ness] to listen to and consider the reasons offered by others' and *rational* if it reflects the pursuit of rational advantage. The distinction, Rawls suggests, reflects Kant's distinction between pure and empirical reason as reflected in the distinction between categorical and hypothetical imperatives (Rawls n.d.: lecture 0, p. 9). The terms reasonable and rational, Rawls suggests, reflect the most fundamental forms of the employment of practical reason and therefore reflect the two fundamental qualities that, in Kant's view, characterize persons as moral beings.

A Kantian conception of the person thus views the person as characterized by four qualities: (i) the rational; (ii) the reasonable; (iii) freedom; and (iv) equality (Rawls 1980: 306). Each of these qualities is represented as a structural feature of the decision procedure. The *rational*, which 'expresses a conception of each participant's rational advantage' (Rawls 1980: 316), is – Rawls claims – straightforward, and is represented simply by the motivations assigned to parties in the original position – their desire to advance their conception of the good and to develop and exercise their moral powers (Rawls 1980: 336). *The reasonable* is defined as the willingness to act from fair terms of cooperation and the ideal of reciprocity and mutuality implicit in those fair terms, and is represented by 'the framework of constraints' within which the parties deliberate (Rawls 1980: 317). The *freedom* of moral persons is represented by several features of the decision procedure, the most significant of which is the fact that the choices of the deliberators are not restricted by any background view limiting permissible conceptions of the good (Rawls 1980: 331). The *equality* of moral persons is represented by two features of the original position: (i) all persons are situated symmetrically with respect to each other – none are assigned superior rights or powers (Rawls 1980: 336); and (ii) the veil of ignorance prevents any person from appealing to superiority in their share of natural or social endowments to justify proposed principles (Rawls 1980: 337–8).

Thus, in Kantian constructivism, the particular principles that give content to Kant's moral view are specified by a decision procedure that is designed to represent the four qualities that constitute the Kantian conception of the person. In modelling these four qualities, the decision procedure thus gives concrete expression to the Kantian intuition that reliable moral judgements are grounded neither in facts nor in independent and pre-existing norms or principles, but rather in a process of reasoning, by establishing the relation that Rawls requires between a Kantian conception of the person and the procedure that defines the content of a Kantian moral view.

3. Constructivism, Facts and Social Practices

In Rawls's account of ethical constructivism, then, the structure of the decision procedure is designed to represent structural features of a process of moral reasoning that is grounded, not in independently existing normative entities or facts, but rather in a process of reasoning. G. A. Cohen and Aaron James, however, argue that the structure represented in that procedure – as that procedure is described in *A Theory of Justice* – *does* assign authority in moral judgement to independently existing normative entities or facts. Both arguments, in fact, attempt to establish that Rawls's ethical constructivism is fact-sensitive *at the foundational level*; and both arguments necessarily fail because the structure of the original position ensures that Rawls's constructivist argument cannot be fact-sensitive at the foundational level. In this section, I will (i) argue that both arguments offer confused readings of Rawls's account of ethical constructivism; and (ii) justify the claim that Rawls's ethical constructivism *cannot* be fact-sensitive at the foundational level.

3.1 Facts and Fact-Dependent Principles

Cohen argues that constructivist approaches in general derive principles of justice from 'considerations of pure non-justice', considerations that include both 'facts about human nature and society' and judgements 'about the right procedure for generating principles of justice' (Cohen 2008: 281). As a result, Cohen claims, constructivism deletes considerations of pure justice – the considerations that should be central to judgements of justice – from the set of factors relevant to the derivation of principles of social justice and, instead, attempts to derive those principles from 'considerations that do not reflect the content of justice' (Cohen 2008: 283).

Rawls's particular account of constructivism, Cohen claims, grounds its arguments in facts both directly and indirectly. That is, Rawls includes

specific facts about human nature and society among the considerations that are considered relevant to the grounding of principles of justice and, in addition, Rawls's arguments appeal to the authority of fact-dependent principles (e.g. the Pareto Principle, the principle of publicity). Because of these aspects of his constructivist method, Cohen concludes, Rawls's account of justice does 'not (really) investigat[e] the nature of justice as such' (Cohen 2008: 301) and 'systematically conflat[es] other questions with the question of justice' (Cohen 2008: 3).

A number of commentators have accepted Cohen's claim that Rawls's constructivist approach grounds principles of justice in facts or fact-dependent principles, while disputing Cohen's claim that fundamental principles of justice *must* be derived from considerations of pure justice (see Barry 1989; Scheffler 2003) Although many of these arguments are persuasive, these commentators nevertheless concede too much to Cohen's critique of Rawls. In fact, the *constructivist* strand of Rawls's argument does *not* ground its principles in facts or fact-dependent principles as Cohen suggests. To the extent that the facts and fact-dependent principles that Cohen cites are relevant to Rawls's justification of his theory, these considerations are employed only in Rawls's *intuitive* argument for the principles of justice, an argument that is separate and entirely distinct from his constructivist argument.[6]

Although Cohen mentions a number of specific facts that he claims ground judgements in Rawls's constructivist argument, he seems most concerned to criticize the role that he claims that *fact-dependent principles* play in Rawls's constructivist approach. Cohen discusses the role of three fact-dependent principles – the Pareto Principle, the principle of stability and the principle of publicity – in Rawls's argument (Cohen 2008: 285–6), and focuses most particularly on the role of the Pareto Principle. In fact, Cohen's concerns regarding the role of the Pareto Principle emerge as a central theme of his final book, *Rescuing Justice and Equality*. Reliance upon arguments grounded in the Pareto Principle, Cohen claims, introduces both theoretical and practical problems into Rawls's argument. As a practical matter, endorsement of the Pareto Principle-based argument that inequalities are acceptable as part of a scheme that makes everyone better off (Cohen calls this argument the 'Pareto argument'; Cohen 2008: 15–16) provides at least indirect support for regressive policies such as the tax cuts of the Thatcher, Reagan and Bush administrations (Cohen 2008: 27–30). In addition, Cohen argues, Rawls's reliance upon the logic of the Pareto Principle generates serious theoretical problems. First, the Pareto argument introduces inconsistency into Rawls's theory, since the morally arbitrary

inequality that is justifiable under the Pareto argument for the difference principle 'contradicts the content of that very principle' (Cohen 2008: 156). Second, reliance upon the Pareto argument renders the case for the difference principle 'incoherent', as Brian Barry's defence of inequality under the difference principle as a 'necessary *evil*' indicates (Cohen 2008: 113; my emphasis).

Remarkably, however, despite Cohen's strenuous criticism of Rawls's reliance upon the Pareto Principle in his constructivist argument, the Pareto Principle plays no role in that argument. In *A Theory of Justice,* Rawls develops two distinct arguments justifying his proposed principles of justice. In chapter 2, Rawls develops an *intuitive* argument for the principles that makes no use of the idea of the original position and argues directly from substantive considered judgements of justice. It might be at least plausible to describe the justification of the difference principle developed in chapter 2 as relying upon the Pareto Principle, but the argument developed in chapter 2 is explicitly not a constructivist argument. In the *constructivist* case for the principles presented in chapter 3, Rawls offers two arguments for the principles, neither of which relies upon the Pareto Principle. In the first argument, Rawls claims that rational choosers in the original position would select his proposed principles because those principles guarantee protections of fundamental liberties that are more satisfactory than those provided under any viable alternative theory (and therefore minimize the strains of commitment). In particular, the principles provide satisfactory protection of liberty interests by securing their *priority* over all other fundamental interests (Rawls [1971] 1999: 154). Second, Rawls argues that the principles guarantee protections of the interest in a *fair* distribution of shares of social goods that are more satisfactory than under alternative theories by 'manifest[ing] in the basic structure of society men's desire to treat one another … as ends in themselves' (Rawls [1971] 1999: 156), thus securing for less advantaged members of society a larger share of the primary good of self-respect than under alternate conceptions of justice (Rawls [1971] 1999: 157–8). It is important to emphasize that neither of these arguments relies upon the Pareto argument. These arguments for the principles offer a *characterization* of the principles' adequacy in addressing fundamental interests and *not* a *description* of their sufficiency in making people better off. Cohen's argument that Rawls's constructivist argument relies upon the Pareto argument therefore fails.

In chapter 2 of *Rescuing Justice and Equality*, Cohen himself appears to concede this point: 'the Pareto argument … is not Rawls's official

argument for the difference principle, since the Pareto argument *dispenses with the device of the original position*' (Cohen 2008: 88; my emphasis). Since Rawls's constructivist argument is *essentially* characterized by its employment of the original position, Cohen's concession that the Pareto argument and the original position are not employed in the same argument is equivalent to a concession that Rawls does not rely upon the Pareto argument in his constructivist argument. In chapter 7, however, Cohen seems not to recall this insight. In that chapter, he simply – and without further argument – lists the Pareto Principle among the three fact-dependent principles that improperly affect Rawls's *constructivist* argument.

Cohen's arguments for the claims that Rawls's constructivist argument is dependent upon the fact-sensitive 'principles' of *publicity* and *stability* are even less persuasive. Cohen's case that Rawls's argument is dependent upon a fact-sensitive principle of publicity fails both because Cohen describes the notion of publicity that is relevant to Rawls's argument incorrectly and because Cohen misunderstands that notion's status in Rawls's argument. Cohen describes the principle of publicity employed in Rawls's argument as requiring 'that it should be possible to tell whether or not someone is observing a principle' (Cohen 2008: 285–6). Rawls, however, presents the idea of publicity as a 'formal constraint of the concept of right', not a principle, and he defines this constraint to require that: (i) 'everyone will know about [the proposed principles of justice] all that he would know if their acceptance were the result of agreement'; and (ii) 'the parties evaluate conceptions of justice as publicly acknowledged' (Rawls [1971] 1999: 115). Once the relevant sense of publicity has been defined correctly, it is clear that publicity does not constitute a ground from which the principles of justice are derived, as Cohen claims. Rather, the formal constraint of publicity merely imposes a condition that the *justification* of the principles must satisfy. In particular, that justification must not rely on 'secret reservation[s]' (Rawls [1971] 1999: 115, n. 8) or deceptions and must present reasons that all reasonable persons could acknowledge and accept. The constraint of publicity thus (i) imposes a merely formal condition on an acceptable justification for a principle of justice and (ii) is not fact-sensitive.[7] Cohen's claim that a fact-dependent principle of publicity improperly grounds Rawls's argument for the principles thus fails.

Cohen defines the fact-sensitive *principle of stability* to require that 'principles governing society should be self-reproducing' (Cohen 2008: 286). Although Rawls does not refer to a *principle* of stability, he does argue that stability is an important consideration in reflection about the

basic requirements of justice. In arguing that a conception of justice must be stable, Rawls means that the principles of justice 'should be such that when they are embodied in the basic structure of society men … develop a desire to act in accordance with its principles' (Rawls [1971] 1999: 119). Rawls does not, however, argue that principles should be designed to secure citizens' *unreflective* acceptance as Cohen seems to suggest. Rather, Rawls's concern once again relates to the kinds of reasons that may ground acceptance of principles of justice. The concern with stability is simply the concern that an acceptable conception of justice must be justified on the basis of reasons that can be accepted by all reasonable persons as reasonable. Once again, the 'principle' that Cohen discusses is merely a formal constraint on the nature of an acceptable justification for an account of justice, and not a substantive and fact-sensitive ground from which the principles are derived. Cohen's argument regarding the principle of stability therefore fails.

What about the specific facts that Cohen argues count among the grounds of Rawls's constructivist argument? Cohen seems most concerned to emphasize the influence of two factual claims on Rawls's argument: (i) well-regulated market economies can function without generating too much inequality; and (ii) people require certain essential goods in order to pursue their life plans (Cohen 2008: 293). Once again, however, Cohen grounds his criticisms in considerations that may plausibly be viewed as relevant to Rawls's intuitive argument, but not to his constructivist argument. As noted above, in the arguments that constitute the 'main grounds' of the *constructivist* case for the two principles, Rawls asserts that the principles: (i) provide protections for liberty interests that are more satisfactory than those provided by any viable alternative principles of justice (by assigning absolute priority to the protection of fundamental liberty interests); and (ii) provide protections for the interest in a *fair* distribution of shares of social goods that are more satisfactory than those provided by any viable alternative principles of justice (by ensuring that society's basic structure 'manifests … men's desire to treat one another … as ends in themselves'; Rawls [1971] 1999: 156). Neither of these arguments depends upon assumptions regarding the functioning of well-regulated market economies or the notion that people require certain essential goods to pursue their life plans effectively.

Rawls's constructivist argument is grounded neither in facts nor in fact-dependent principles. Rather, that argument focuses narrowly on the question of what principles free and equal rational persons, choosing

from a fair standpoint, would view as providing satisfactory protections for their fundamental interests. This specific focus reflects the influence on Rawls of the Kantian intuition that the substance of morality is best understood as constructed by free and equal people under fair conditions. Cohen's confusion regarding the character of Rawls's argument thus reflects a failure to recognize the central role of this Kantian intuition in Rawls's theory.

3.2 Existing Social Practices

Like Cohen, James argues that Rawls's account of constructivism does not succeed in providing an account of moral and political judgement that is grounded in *a process of reasoning* rather than in facts or in fact-dependent premises. James argues that Rawls's account of constructivism in fact grounds the authority of moral reasoning in independent judgements about the nature and point of existing social practices. The authority of judgements regarding the purposes of distinct social practices in Rawls's account of moral reasoning, James argues, helps to explain why Rawls's account of domestic justice focuses on what is owed to *persons* while his account of justice in the global setting focuses on what is owed to *peoples*. In each setting, James argues, Rawls grounds his reasoning in judgements about the nature of existing practices and the agents who participate in them.

Rawls's constructivist method, James suggests, has been fundamentally misunderstood. Rather than working from free-standing considered judgements of justice to the description of a standpoint from which moral reasoning has independent and absolute authority, Rawls has instead 'been following a single, abstract "constructive" method, which begins from existing social practices' (James 2005: 282). Rawls's method has involved identifying social practices and their participants, specifying the practice's point or goal, representing participants as appropriately motivated by 'an interest in the goods that the practice is designed to create', and designing a veil of ignorance that provides *all* parties with the same information but *no* parties with information that could undermine the fairness of any agreements reached (James 2005: 282). Once we recognize that this description fits Rawls's approach, James notes, it will be clear that 'original position reasoning has no authority as such; it must be grounded in independent judgments about what practices exist and what kinds of agents participate in them' (James 2005: 282). Moreover, since Rawls – in James's view – offers his theory as 'an interpretive characterization of our basic structures', his arguments cannot properly make 'any direct appeal to moral considerations' (James 2005: 305).

While James's account of constructivism as the constructive inter-pretation of social practices is of some interest on its own merits, that account cannot plausibly be described as an interpretation of Rawls's constructivism. Three claims are central to James's interpretation of Rawls's approach: (i) Rawls's constructivism limits the grounds of moral and political reflection to independent judgements about existing social practices; (ii) Rawls's approach can make no direct appeals to moral considerations; and (iii) original position reasoning possesses no authority independent of judgements about existing practices and participants in those practices. None of these claims, I will argue, is consistent with Rawls's account of his approach to constructivism.

Arguments from existing practices. James is correct to note that persons engaged in reflection regarding moral and political questions will, in Rawls's account, include among the considerations most fundamental to their deliberations considered judgements regarding the nature and internal logic of their social practices. While reflection does *begin* with a set of ideas that focuses on such practices, however, Rawls does not argue that reflection should be limited to arguments grounded in this set, nor is reflection limited to appeals to *any* set of judgements relating to the background political culture.[8] Persons reflecting on moral and political questions, Rawls argues, properly *assess and revise* any and all notions received from their tradition.[9] Each person 'look[s] to the public culture as the shared fund of implicitly recognized basic ideas and principles'; *but* in order 'to be acceptable', our judgements of justice must accord, not with some description or sympathetic reconstruction of the nature and goals of shared practices, but 'with our considered convictions, at all levels of generality, on due reflection … [in] "reflective equilibrium"' (Rawls [1993] 1996: 8). And in due reflection, Rawls argues that the person must consider '*all possible descriptions* to which one might plausibly conform one's judgments' (Rawls [1971] 1999: 43, my emphasis), and should take note of 'alternate conceptions of justice and the force of various arguments for them' (Rawls [1993] 1996: 384, n. 6). Rawls thus explicitly rejects the claim that the grounds of moral and political reasoning are limited to information regarding existing practices and their participants.

Appeals to moral considerations. James claims that Rawls's con-structivism precludes direct appeals to moral considerations in grounding moral or political judgements. This claim, however, confuses the terms of debate in the original position with the standards of justification in Rawls's constructivism. It is certainly true that Rawls rules out moral considera-tions as grounds for judgements *by the choosers in the original position*. The choosers 'decide solely on the basis of what seems best calculated to

further their interests' (Rawls [1971] 1999: 512). But this limitation on grounds for judgement merely reflects a division of labour designed to clarify the functions of various elements of Rawls's decision procedure: '[i]t is … to preserve [clarity between the differing relevant ethical considerations relevant to choice] that I have avoided attributing to the parties [in the original position] any ethical motivation' (Rawls [1971] 1999: 512). The original position *itself* is in fact designed to represent fundamental *moral* considered judgements regarding the conditions under which judgements about fundamental questions of justice should be formed. The veil of ignorance represents the considered judgements that (i) considerations that are 'irrelevant from the standpoint of justice' (Rawls [1971] 1999: 17) should not determine the distribution of social goods and (ii) 'it should be impossible to tailor principles of justice to the circumstances of one's own case' (Rawls [1971] 1999: 16); the veto that each chooser possesses over any proposed principle represents the considered judgement that justice must respect the inviolability of the person (Rawls [1971] 1999: 3); and the symmetrical situation of the choosers represents the considered judgement that the substance of morality is best understood as constructed by free and equal people under fair conditions. As Rawls notes, the design of the original position 'includes moral features and must do so' (Rawls [1971] 1999: 512). While the judgements of the choosers in the original position do not appeal to moral considerations, then, the judgements that ground acceptance of the original position as the preferred standpoint from which to judge questions of justice appeal directly to moral considerations, and judgements that are the product of original position reasoning are therefore grounded directly in moral propositions that are embodied in the original position's structure and not merely in norms associated with existing practices. James's argument that Rawls's constructivism precludes direct appeals to moral considerations therefore fails.

Authority independent of judgements about existing practices. James argues that original position reasoning possesses no authority independent of judgements about existing practices and participants in those practices. This claim, however, assumes the correctness of James's argument that Rawls's constructivism limits the grounds of moral and political reflection to independent judgements about existing social practices. If, as I have argued above, the grounds of moral judgement in Rawls cannot plausibly be viewed as limited to independent judgements about existing social practices, then the grounds of moral judgement constitute the entire set of considered judgements relevant to justice, and the authority of many judgements formed in the original position derive from their grounding in relevant considerations contained in this set that are not judgements about

social practices. Note, moreover, that Rawls's argument for the authority of the Law of Peoples demonstrates that Rawls must reject James's claim about the limited authority of original position reasoning, since Rawls argues that (i) judgements formed in a suitably tailored original position should be viewed as authoritative trans-culturally by all reasonable or decent peoples; and (ii) Rawls assumes that the members of the set of reasonable or decent peoples do not share common practices. If the authority of original position reasoning were limited as James suggests, the Law of Peoples could not exercise authority for more than a single people. Thus, James's argument regarding the authority of original position reasoning is clearly inconsistent with Rawls's understanding of the scope of the authority of original position reasoning.

Conclusion.   I want to conclude this section by justifying the claim that the arguments offered by Cohen and James *must* fail, because justifications generated from the standpoint of the original position cannot be dependent on facts or fact-sensitive principles at the foundational level. As discussed above, the structure of the original position is designed to embody four foundational considered judgements: (i) 'it should be impossible to tailor principles of justice to the circumstances of one's own case' (Rawls [1971] 1999: 16); (ii) principles of justice may not be justified on the basis of considerations that are 'irrelevant from the standpoint of justice' (Rawls [1971] 1999: 17) or 'arbitrary from the moral point of view' (Rawls [1971] 1999: 63); (iii) each person possesses an inviolability founded on justice (Rawls [1971] 1999: 3); and (iv) the substance of morality is best understood as constructed by free and equal people under fair conditions. Four *fact-insensitive* principles associated with these considered judgements – (i) justice must be impartial; (ii) justice must not be arbitrary, (iii) justice must respect the inviolability of the person; and (iv) principles of justice should correspond to the principles that would be chosen by reasonable and rational persons under conditions that characterize them as free and equal – are embodied in structural features of the original position (e.g. the veil of ignorance, the symmetrical positions of the choosers, the veto that each chooser may exercise over proposed principles) and ground every argument made from that standpoint. If an argument can be made or a judgement can be reached in the original position, it is because that argument or judgement constitutes a practical implication or extension of one or more of these foundational and *fact-independent* principles.

Cohen argues that a principle can respond to a fact only because it is also a response to a principle that is not a fact or a fact-sensitive

principle. But this is precisely the case in the original position – any consideration that is relevant to the justification of principles of justice in the original position is relevant precisely because at least one of the four foundational fact-independent principles embodied in its structure justifies the view that that consideration is relevant; and if the principles of justice respond to any facts in the original position, that response is also a response to one of the fact-independent principles embodied in the structure of the original position.

Facts, fact-sensitive principles and practices may be relevant to the *intuitive* argument for the principles of justice, but the structure of the original position ensures that the justification of the constructivist argument cannot be dependent upon such facts, fact-sensitive principles or practices at the foundational level. Since both Cohen and James offer interpretations of the *constructivist* argument, their claims that Rawls's argument is dependent at the foundational level on facts, fact-sensitive principles and/or practices therefore *necessarily* fail.

4. Kantian Objections

While Rawls argues that the central role that his account of constructivism assigns to a decision procedure constitutes a procedural interpretation of Kant's ideas regarding moral reasoning and autonomy, a number of commentators – in particular, Larry Krasnoff (1999) and Onora O'Neill (1989) – have raised important objections to this claim. In particular, these commentators argue, Kant's account of moral reasoning employing the categorical imperative fails to 'fit well' (Krasnoff 1999: 401) with the central features of Rawls's account of constructivism. First, the categorical imperative does not enable moral reasoners to generate moral principles; rather, it serves as a 'negative check' on the maxims of action that individuals assess in moral deliberation. Second, Rawls's decision procedure imposes more stringent restrictions on the scope of ethical concern than Kant's procedure requires. Finally, the categorical imperative does not model hypothetical choice; rather, it tests whether all persons could *possibly* choose to act from a particular maxim or principle of volition.

These objections, I will argue, fail. While Rawls's critics in fact exaggerate the distinctions between the two decision procedures, these objections are unpersuasive primarily because they argue from the incorrect assumption that Rawls presents his decision procedure as equivalent to the categorical imperative procedure. Rawls makes no such claim – his account of constructivism is Kantian, he argues, because it develops Kant's view that

the substance of morality is best understood as constructed by free and equal people under fair conditions.

4.1 The Choice of Principles?

While the categorical imperative serves merely as a negative check on the person's specific maxims or principles of volition, Krasnoff (1999: 400) and O'Neill (1989) argue, the original position is designed to enable the parties to generate new and general moral principles.[10] Rawls's decision procedure thus performs a task that is both more ambitious and less specifically practical than that performed by the categorical imperative. The output of Rawls's procedure is, these critics conclude, distinguishable from that of the categorical imperative procedure in two important respects. Principles generated in the original position (i) apply to a broader range of problems than maxims endorsed through the categorical imperative procedure and (ii) are newly constructed, rather than merely endorsed from a rational standpoint.

First, it is important to note that this objection exaggerates the distinctions between the two procedures. While the categorical imperative procedure generally focuses on specific practical maxims that require specific acts or omissions, its proper employment is not limited to such maxims. The most common instances analysed under Kant's procedure involve maxims connected with specific moral choices, but Kant also discusses cases in which application of the categorical imperative procedure requires that the person adopt wide duties (duties to adopt an end). Thus, for example, Kant discusses applications of the categorical imperative requiring that the moral person must recognize wide duties to develop her talents and to adopt the happiness of others as her own end (Kant 1996: 81; G 4: 430; Kant 1996: 518; DV 6: 387). These wide duties cannot, by definition, be satisfied by specific acts; rather, they require that the agent adopt general policies designed to further the respective ends of one's own perfection and the happiness of others. Rawls's decision procedure, then, does not necessarily focus on problems that are different in character from those examined by moral reasoners employing the categorical imperative procedure.

In addition, deliberations in the original position do not – as this objection suggests – generate fresh and newly minted moral principles. Rather, just as individuals applying the categorical imperative procedure consider and assess candidate maxims and principles of volition, parties in the original position consider and assess candidate principles. Rawls discusses 'the presentation of alternatives' explicitly in chapter 3

of *A Theory of Justice* (Rawls [1971] 1999: 105–9). Ideally, Rawls states, 'one would like to say that [the parties] are to choose among all possible conceptions of justice' (Rawls [1971] 1999: 105). For the purposes of the exposition of his argument, however, Rawls 'resort[s] to the … device' (Rawls [1971] 1999: 105) of listing the leading traditional views for consideration by the parties. While not exhaustive of all theoretical possibilities, the list includes utilitarian, perfectionist, intuitionist, egoist and mixed (utilitarian and liberal) conceptions, as well as justice as fairness. The parties choose from among the sets of principles that define these conceptions. An argument for any candidate set of principles, Rawls argues, 'is always relative to some list of alternatives' (Rawls [1971] 1999: 109), and the best way to identify alternatives that are of 'philosophical interest' is to examine traditional conceptions.

Most significantly, however, Rawls does not claim that the original position is formally or substantively equivalent to the categorical imperative procedure. Rather, Rawls argues that his procedure represents the 'clearly discernible' *character* of Kant's moral theory by 'enabl[ing] us to explain the sense in which acting from principles expresses our nature as free and equal rational persons' (Rawls [1971] 1999: 226–7). As discussed in section 2, Rawls argues that it is not the structure of the procedure itself, but rather the relation between that structure and a Kantian conception of the person, that is most fundamental to Kantian constructivism. A Kantian conception of the person views the person as characterized by four qualities (the rational, the reasonable, freedom, equality), and each of these qualities is represented as a structural feature of the decision procedure.

Rawls, moreover, explicitly contrasts the distinctly social focus of his employment of Kantian constructivism in *A Theory of Justice* with the ethical focus of the categorical imperative. Kant, Rawls notes, 'proceeds from the particular, even personal, case of everyday life; he assumed that this process carried out correctly would eventually yield a coherent and sufficiently complete system of principles'. Rawls's employment of Kantian constructivism in *A Theory of Justice*, he concedes, 'moves in quite the reverse pattern', aiming directly to identify acceptable principles to regulate the basic structure of a just society (Rawls 1980: 339). By modelling a Kantian conception of the person while reversing the direction of Kant's analysis, Rawls hopes to avoid the notorious difficulties associated with Kant's attempts to specify principles of social justice and an ideal of social life.[11]

4.2 The Scope of Ethical Concern

O'Neill, however, objects that it is precisely the social focus of Rawls's theory that distinguishes his approach to moral reasoning from Kant's moral thought. While Kant's constructivism conceives of human beings as merely 'a plurality of agents lacking antecedent principles of coordination', Rawls's Kant*ian* constructivism presents moral reasoners as citizens embedded within a particular political culture. The narrower focus of Rawls's constructivism restricts 'the scope of ethical concern' to 'a bounded liberal society'. Thus while Kant's moral theory is cosmopolitan, Rawls's theory is 'implicitly statist' (O'Neill 2008: 362).

It is important, however, to distinguish Rawls's argument for his conception of political justice from his account of Kantian constructivism. Kantian constructivism constitutes an approach to moral reasoning that defines a standard of justification that Rawls's theory of justice must itself satisfy. As Rawls emphasizes, his employment of Kantian constructivism in *A Theory of Justice* constitutes merely 'one … variant' (Rawls 1980: 303) of that approach, a variant designed to focus moral judgement on questions of justice. Similarly, political liberalism constitutes merely a possible output of his constructivist approach to moral reasoning. It is in Rawls's arguments for political liberalism, and not in his general account of Kantian constructivism, that citizens are presented as embedded within a particular political culture. Political liberalism's narrower focus on the ethical concerns of 'a bounded liberal society', then, does not represent a feature of Kantian constructivism as a method.

Moreover, in assuming that a form of decision procedure specifically designed to address social questions may be appropriate for examining the subset of moral issues relating to matters of social justice, Rawls follows Kant's approach to moral reasoning as O'Neill herself understands it. As O'Neill points out in *Acting on Principle*, Kant requires different forms of decision procedure to address different kinds of ethical problems, requiring different forms of the categorical imperative procedure, for example, to determine an act's moral and legal status (O'Neill 1975: 74–5). O'Neill is at least potentially inconsistent, then, in arguing that Rawls's account of Kantian constructivism is unfaithful to Kant simply because Rawls's decision procedure is tailored to apply specifically to a distinctly social subset of moral issues.

Finally, it is important to note again that Rawls presents his decision procedure as a faithful interpretation of Kant's ethical thought, not

because that decision procedure is equivalent to the categorical imperative procedure, but rather because it develops faithfully Kant's notion that the substance of morality is best understood as constructed by free and equal persons under fair conditions. Thus, O'Neill's objection, like the arguments of Cohen and James discussed in section 3, reflects a failure to appreciate the central significance in Rawls's argument of this Kantian intuition. Rawls's procedure constitutes an interpretation of this view because (i) the principles chosen are those that would be chosen by reasonable and rational persons under conditions that characterize them as free and equal; (ii) the description of procedure 'enables us to explain the sense in which acting from these principles expresses our nature as free and equal' (Rawls [1971] 1999: 226); and (iii) the employment of the procedure connects Kant's basic ideas concretely with human conduct.[12] The Kantian character of Rawls's constructivist procedure, as Rawls emphasizes, thus derives from its representation of 'a certain conception of the person as free and equal [and] as capable of acting both reasonably and rationally' (Rawls 1980: 306), and not from the superficial similarity of the original position and the categorical imperative procedure.

4.3 Hypothetical or Possible Choice?

While Rawls's decision procedure asks what principles of justice could secure the hypothetical consent of rational and independent choosers, O'Neill argues, Kant's procedure asks whether all persons *could possibly* consent to proposed principles (O'Neill 1989: 110–11, 216–17). Krasnoff, who discusses and develops O'Neill's objection, therefore argues that a fundamental asymmetry exists between the original position and the categorical imperative procedure (Krasnoff 1999: 401). Rawls's procedure aims to identify a unique set of principles that would be the hypothetical choice of a designated set of deliberators, while Kant's procedure asks whether a particular principle could possibly be accepted by all persons who are actually to be subject to the authority of those principles.

This distinction is significant, O'Neill argues, because a theory employing a standard of hypothetical choice seems designed to identify principles that represent an equilibrium among the competing subjective interests of the choosing parties. A hypothetical agreement among the parties securing such an equilibrium, O'Neill claims, would simply represent the judgement that the principles chosen secured a balance of interests among the parties more effectively than the alternate candidates considered. O'Neill is particularly concerned that a theory employing a

hypothetical consent standard would: (i) justify the coercive enforcement of principles against people who do not *actually* consent; and (ii) allow the preferences and desires of the parties to play an inappropriate role in determining the content of a theory of justice (O'Neill 1989: 109). A theory employing a standard of possible choice, in contrast, aims merely to avoid the adoption of principles or maxims that could not be accepted by persons who will be affected by the adoption of the principle or maxim. As O'Neill notes, Kant's argument against deceit, in the *Groundwork*, provides a paradigmatic example of a maxim to which an affected person could not consent (O'Neill 1989: 112–13). If I act on a maxim of misleading a person in order to use her for my purposes, the person 'cannot possibly agree with my way of behaving to him' (Kant 1996: 80; *G* 4: 429). Under such a possible consent approach, both of O'Neill's concerns are addressed. First, principles to which others do not actually consent will not be adopted under a possible consent standard; and second, fundamental interests, rather than preferences and tastes, will determine the principles adopted under such an approach.

Once again, however, Rawls's critics exaggerate the significance of the distinction between these two standards. Persons deliberating about justice who adopt the perspective of the original position in order to focus their reflections are not aiming to secure equilibrium among competing subjective interests. Rather, such persons – in reflecting subject to the formal and substantive constraints that are imposed upon original position deliberations – are most accurately described as assessing the *possible* choice of the principles under consideration. The original position constraints – including the symmetrical position of the parties, the impossibility of appealing to the person's own situated interests, and the lack of information about the deliberator's own conception of the good – represent an undertaking to choose principles in the spirit of mutual respect and reciprocity. In choosing principles subject to these constraints, then, each deliberator may be viewed as asking herself: could the other members of society who will be affected by the principles under consideration *possibly* consent to their adoption? Thus, it is plausible to characterize deliberations in the original position, as well as deliberations employing the categorical imperative procedure, as testing proposed principles against a *possible consent* standard.

In fact, an argument of Thomas Hill's suggests that this conclusion should not surprise us because the distinction between possible and hypothetical consent 'is not in itself deeply significant' (Hill 2002: 65). The standard of rationality that guides judgements regarding hypothetical

consent, Hill argues, can be reformulated in terms of what rational agents could possibly consent to.

> The rational standards on which the 'could will' test relies can … be expressed in terms of what rational agents necessarily 'would will if rational.' Moreover … the prohibitions that hypothetical rational agents would will are just those that are rationally necessary for them to will, given their situation. Both formulas, then, presuppose as background some general standards of rational willing. (Hill 2002: 65)

Rawls's standard of hypothetical consent does not, Hill concludes, necessarily perform a role in moral reasoning materially different from or inconsistent with the role performed by Kant's standard of possible consent.

Conclusion: Constructivism and Justification

Rawls's account of Kantian constructivism develops and extends the implications of Kant's view that moral judgements are grounded neither in an independently existing order of values nor in special features of human psychology, but rather in a process of reasoning. Kantian intuitions thus lead Rawls to focus his moral analysis on the notion of a theory of social contract designed to secure reasonable and mutually justifiable social relations, and Rawls's development of the doctrine of social contract grounds his accounts of Kantian constructivism, reflective equilibrium and justice as fairness.

In his later work on political liberalism, Rawls modifies his account of the form of constructivism that reasonable and rational persons would accept in reflective equilibrium. Rawls concludes that such persons would judge that *political* constructivism, rather than Kantian constructivism, constitutes the most acceptable method to determine the choice of principles to regulate their judgements of political justice. Rawls modifies his account of constructivism because of two concerns. First, Rawls concludes that members of a society characterized by reasonable pluralism would not generally accept the Kantian conception of the person as the foundation of their moral judgements. The second concern relates to the status of judgements grounded in a constructivist approach. While political constructivism asserts that principles of political justice may be *represented* as the product of a decision procedure, this approach emphatically avoids the claim that such principles *must* be constituted through such a procedure. This emphasis is designed to address the concern that various conceptions of the good may (i) hold that

principles of justice and morality are true; and (ii) state explicit criteria to determine the truth of claims of justice and morality. While persons holding such views would necessarily reject a form of constructivism that claimed that moral and political principles *must* be the product of a constructivist approach, such persons could, Rawls argues, accept a form of constructivism that claimed merely that moral and political principles *may be represented* as the product of such an approach.

Ideas from Kant's political thought, however, continue to exert a powerful influence on Rawls's later work. Ideas from Kant are, in fact, constitutive of his account of *political* constructivism in two important respects. First, political constructivism preserves the emphasis on reasoned agreement that Rawls derives from Kant's theory of social contract. And second, as discussed above, the standard of reflective equilibrium upon which Rawls relies to justify the authority of constructivist reasoning is grounded – in significant part – in Kantian intuitions. Finally, *Kantian* constructivism – as an approach distinct from political constructivism and independent of Rawls's substantive account of justice – remains of interest both because the theory (i) sets out a distinctive non-consequentialist account of justification in ethics and (ii) continues to influence a wide range of moral philosophers.

E-mail: akaufman@uga.edu

Notes

1. Samuel Freeman provides a helpful discussion of intellectual influences on Rawls's account of justification in *Rawls* (2007: 23–5).
2. This view of ethics, Rawls claims, is 'perfectly clear in its Kantian formulation' (Rawls 1967: 132).
3. 'Kant's idea of autonomy requires that there exists no [independent] order of objects determining the moral duties among free and equal persons' (Rawls n.d.: lecture 0, p. 24). 'Kant insists that a good will is not good because of … its fitness to some independently specified … end' (ibid.: lecture 2, p. 7).
4. References to and citations of Kant's work are given parenthetically in the text using the following abbreviations and citing the page numbers of the relevant volume of *Kants gesammelte Schriften* (published by the Preussische Akademie der Wissenschaften, Berlin). *Grounding of the Metaphysics of Morals* (G) (1785); The Doctrine of Virtue *(Tugendlehre)* (DV) (1797); The Doctrine of Right (*Rechtslehre*) (R) (1797); 'On the Common Saying: "This May Be True in Theory, But it does Not Apply in Practice"' (TP) (1793). I have used the translations of these works contained in *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*, trans. Mary J. Gregor (Cambridge: Cambridge University Press, 1996).
5. Rawls 1967: 131.
6. Rawls's intuitive argument, presented in ([1971] 1999: ch. 2), develops a justification for the second principle of justice directly from substantive considered judgements

of justice. This argument is not constructivist and makes no use of the original position.

7   Rawls, in fact, asserts that '[t]he publicity condition is clearly implicit in Kant's doctrine of the categorical imperative' (Rawls [1971] 1999: 115).

8   '[T]here are no judgments on any level of generality that are in principle immune to revision' (Rawls 1975: 289).

9   '[W]e may [on reflection] reaffirm our particular judgments and decide instead to modify the proposed conception of justice with its principles and ideals … It is a mistake to think of abstract principles and general conceptions as always overriding our more particular judgments' (Rawls [1993] 1996: 45).

10   'Kantian maxims do not entail … rules for all possible contexts' (O'Neill 1989: 117).

11   See Kaufman (1999) for an account of political judgement in Kant's later political writings.

12   'No longer are these notions purely transcendent and lacking explicable connections with human conduct' (Rawls [1971] 1999: 226).

References

Barry, Brian (1989) *Theories of Justice*. London: Harvester Wheatsheaf.

Cohen, G. A. (2008) *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.

Freeman, Samuel (2007) *Rawls*. London: Routledge.

Hill, Jr., Thomas E. (2002) *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford: Oxford University Press.

James, Aaron (2005) 'Constructing Justice for Existing Practice: Rawls and the Status Quo'. *Philosophy and Public Affairs*, 33/3, 281–316.

—— (2007) 'Constructivism about Practical Reasons'. *Philosophy and Phenomenological Research*, 74/2, 302–25.

Kant, I. (1996) *Practical Philosophy*. Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press.

Kaufman, Alexander (1999) *Welfare in the Kantian State*. Oxford: Oxford University Press.

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

—— (2008) *The Constitution of Moral Agency*. Oxford: Oxford University Press.

Krasnoff, Larry (1999) 'How Kantian is Constructivism?'. *Kant-Studien*, 90, 385–409.

O'Neill, Onora (1975) *Acting On Principle: An Essay in Kantian Ethics*. New York: Columbia University Press.

—— (1989) *Constructions of Reason*. Cambridge: Cambridge University Press.

—— (2008) 'Constructivism in Rawls and Kant'. In Samuel Freeman (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press), pp. 347–67.

Rawls, John (1951) 'Outline of a Decision Procedure for Ethics'. *Philosophical Review*, 60/2, 177–97, Repr. in Rawls 1999: 1–19.

—— (1955) 'Two Concepts of Rules'. *Philosophical Review*, 64/1, 3–2, Repr. in Rawls 1999: 20–46.

—— (1958) 'Justice as Fairness'. *Philosophical Review*, 67/2, 164–94, Repr. in Rawls 1999: 47–72.

—— (1963a) 'Constitutional Liberty and the Concept of Justice'. In *Nomos*, vol. 6, *Justice*. New York: Atherton Press), pp. 98–125, Repr. in Rawls 1999: 73–95.

—— (1963b) 'The Sense of Justice'. *Philosophical Review*, 72/3, 281–305, Repr. in Rawls 1999: 96–116.

—— (1964) 'Legal Obligation and the Duty of Fair Play'. In Sidney Hook (ed.), *Law and Philosophy: A Symposium* (New York: New York University Press), pp. 3–18. Repr. in Rawls 1999: 117–29.

—— (1967) 'Distributive Justice'. In Peter Laslett and W. G. Runciman (eds), *Politics, Philosophy, and Society*. Oxford: Blackwell), Repr. in Rawls 1999: 130–53.

—— ([1971] 1999) *A Theory of Justice*. Cambridge, MA: Harvard University Press.

—— (1971) 'Justice as Reciprocity'. In S. Gorowitz (ed.), *John Stuart Mill: Utilitarianism, with Critical Essays*. Indianapolis, IN: Bobbs Merrill), Repr. in Rawls 1999: 190–224.

—— (1975) 'The Independence of Moral Theory'. *Proceedings and Addresses of the American Philosophical Association*, 48, 5–22, Repr. in Rawls 1999: 286–302.

—— (1980) 'Kantian Constructivism in Moral Theory'. *Journal of Philosophy*, 77, 515–72, Repr. in Rawls 1999: 303–58.

—— (1999) *Collected Papers*, ed. Samuel Freeman Cambridge, MA: Harvard University Press.

—— (n.d.) 'Lectures on Kant's Ethics'. Unpublished manuscript.

Scanlon, T. M. (1998) *What we Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scheffler, Samuel (2003) 'Rawls and Utilitarianism'. In Samuel Freeman (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press), pp. 426–59.