https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

THE COMMONWEALTH OF BEES: ON THE IMPOSSIBILITY OF JUSTICE-THROUGH-ETHOS*

By Gerald Gaus

Abstract: *Some understand utopia as an ideal society in which everyone would be thoroughly informed by a moral ethos: all would always act on their pure conscientious judgments about justice, and so it would never be necessary to provide incentives for them to act as justice requires. In this essay I argue that such a society is impossible. A society of purely conscientiously just agents would be unable to achieve real justice. This is the Paradox of Pure Conscientiousness. This paradox, I argue, can only be overcome when individuals are prepared to depart from their own pure, conscientious, judgments of justice.*

KEY WORDS: ideal theory, egalitarian ethos, incentives, public justice, public reason, social rules

> It is true, that certain living creatures (as bees, and ants), live sociably one with another, (which are therefore by *Aristotle* numbered amongst political creatures), and yet have no other direction, than their particular judgements and appetites, nor speech, whereby one of them can signify to another, what he thinks expedient for the common benefit; and therefore some man may perhaps desire to know, why mankind cannot do the same.
>
> — Hobbes, *Leviathan*

I. A Harmonious World of Perfectly Just Agents?

It is tempting to understand "nonideal" political theory as that which accommodates human frailties such as selfishness or narrow loyalties, while an "ideal" or "utopian" theory is one that depicts a society of perfect moral agents.[1] In such an ideal society everyone would be thoroughly

* My thanks to the other contributors to this volume for their valuable comments and suggestions. Special thanks for the detailed and extremely helpful comments of Cheshire Calhoun, Michael Huemer, and David Wiens.

[1] I do not use "utopian" in a pejorative way, or as an indication that the specified social arrangement is infeasible. Some contemporary political philosophers explicitly contrast "utopian" with "realistic" theory (Laura Valentini, "Ideal vs. Non-ideal Theory: A Conceptual Map," *Philosophy Compass* 7, no. 9 [2012]: 654–64, at 654; see also Robert Jubb, "Tragedies of Non-ideal Theory," *European Journal of Political Theory* 11 [2012]: 229–46, at 230). However, traditional utopian thought was often concerned with viable institutional schemes. (See Barbara Goodwin and Keith Taylor, *The Politics of Utopia* [London: Hutchinson, 1982], 210–14; Timothy Kenyon, "Utopia In Reality: 'Ideal' Societies in Social and Political Theory," *History of Political Thought* 3 [January 1982]: 123–55.) As I use it, a utopian theory is a type of "ideal theory," which may well be implementable. Karl Kautsky famously praised More's *Utopia* as articulating a socialist ideal that satisfied important "realization" constraints. Karl Kautsky, *Thomas More and His Utopia* (London: Lawrence and Wisehart, 1979 [1888]), 249.

doi:10.1017/S0265052516000352

© 2016 Social Philosophy & Policy Foundation. Printed in the USA.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

informed by a moral ethos: all would always act on their pure conscientious judgments about justice, and so it would never be necessary to provide inducements for them to act as true justice requires. As is well known, John Rawls advanced a principle of distributive justice — his so-called "difference principle" — that allows departures from strict egalitarian distributions of resources when those departures provide incentives to the better-off to perform tasks that ultimately contribute to raising the life prospects of those who are least well off.[2] G. A. Cohen finds the need for such incentives inconsistent with an unblemished commitment to justice. He depicts their need as an implicit threat that the better-off will go on strike if it is demanded that they accord full egalitarian justice to the least well off, and even suggests a parallel to a kidnapper's threat that the parents can become better off (get their child back) if ransom demands are paid.[3]

And so an attractive idea suggests itself: if a society was entirely, thoroughly, devoted to egalitarian justice, it could do away with the need for what seems to be grubby, very much nonideal, incentives. If each person's deliberations are thoroughly infused with a devotion to egalitarian justice, the ideal of a perfectly harmonious egalitarian order seems achievable — each freely does as she concludes egalitarianism requires. In a utopian society justice would be secured without incentives. Thus, in Cohen's toy model of a socialist community as a camping trip, people take on the "spirit" of the trip, and desire to serve others while being served by them, thereby generating a spontaneous blending of free choices.[4] As Jason Brennan observes, for Cohen "[a] fully just society is . . . a society in which every person always does the right thing for the right reasons."[5] The ideal of a harmonious blending of free, fully conscientious moral, choices has a long and deep history in utopian thinking, especially nineteenth-century socialist utopianism, from the somewhat wacky Charles Fourier, through Robert Owen and Peter Kropotkin, to utopian novelists such as Edward Bellamy.[6] Under the correct communal ethos and education, the conscientious choices of each would be harmonized — recall that the most famous Owenite socialist community was "New Harmony."

In this essay I argue that such a society is impossible. A society of purely conscientiously just agents would be unable to achieve justice. I shall call this the Paradox of Pure Conscientiousness. This paradox, I shall argue,

[2] John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: Harvard University Press, 1999), 65–73.

[3] G. A. Cohen, *Rescuing Justice and Equality* (Cambridge, MA: Harvard University Press, 2008), 33, 48. "The incentives argument has something in common with the kidnapper argument . . . " (ibid., 41).

[4] G. A. Cohen, *Why Socialism?* (Princeton, NJ: Princeton University Press, 2009), chap. 2.

[5] Jason Brennan, *Why Not Capitalism?* (New York: Routledge, 2014), 71.

[6] See Charles Fourier, *Design for Utopia: Selected Writings of Charles Fourier*, trans. Julia Franklin (New York: Schocken, 1988); Robert Owen, *A New View of Society, and Other Writings* (London: Dent, 1972 [1813]); Edward Bellamy, *Looking Backward* (New York: Dover, 1996 [1880]); Peter Alekseevich Kropotkin, *Mutual Aid: A Factor of Evolution* (London: W. Heinemann, 1907).

can only be overcome when individuals are prepared to depart from their own pure, conscientious, judgments of justice, by grounding their judgments and actions on other values — values that constitute incentives for departing from their purely conscientious judgments.

II. The Paradox of Pure Conscientiousness

A. *Purely conscientious judgment*

Suppose that Alf is a competent moral reasoner in utopia. I assume this implies a high level of moral excellence, for we should expect the competency standards of utopia to be high. A common view of moral reasoning is:

> ASSUMPTION 1: THE "I CONCLUDE WE OUGHT" VIEW. As a competent member of utopia, if (i) Alf conscientiously deliberates and concludes that, given what he takes to be the correct view of justice and relevant empirical information, ϕ is required by justice under conditions *C*, where (ii) this does not require accommodating the conclusions of the deliberations of others, (iii) then he reasonably concludes that justice instructs that we all ought to ϕ under conditions *C*, and so (iv) he justifiably ϕs in circumstances *C* and demands that others do so as well.

The "I conclude we ought" View, which I think is characteristic of much moral philosophy, understands reasoning about moral claims to be, in a fundamental sense, akin to reasoning about ordinary factual claims. On this commonsense approach, when Alf deliberates about a matter of justice, he considers the best reasons as he understands them, including what he takes to be the correct normative principles, *perhaps* checks his conclusions with others to see if he has made any errors, and then comes to the conclusion, "we all ought to ϕ." His moral reasoning may refer to facts about other people (say, their welfare), but it is not a general requirement on the moral reasoning of any competent agent that he accommodates his conclusions to those of others.[7] A purely conscientious judgment, then, is one that is the conclusion of "I conclude we ought" reasoning by Alf, our excellent moral reasoner; a purely conscientious agent is one who always acts on such judgments. Let us call these "private judgments of justice": an agent who always acts on his private judgments is a "resolute moralist."

B. *A community of conscientious agents*

Alf, then, employs his "I conclude we ought" reasoning to arrive at judgments of justice; to fix ideas, I stipulate throughout that he and all others in his community endorse an egalitarian conception of justice. Now if he is

[7] I consider the effects of relaxing (*iii*), the non-accommodation clause, in Section II.B.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

thoroughly informed by an ethos of justice, he will always act in accordance with his moral judgment. He is, let us say, characterized by

> ASSUMPTION 2: THE JUSTICE COMMITMENT. As a competent member of utopia, Alf will always act in accordance with what he deems required by egalitarian justice.

If Alf deems that ϕ is, all things considered, required by justice, then he will ϕ; he will refrain from acting on alternatives that he thinks are flawed from the perspective of justice. "If I conscientiously believe that an act is required by justice," Alf reasons, "then I must do it."

At this point we have not fully captured the ethos of the egalitarian camping trip, which sketches an egalitarian *community* in which excellent moral agents freely act on their private judgments about justice. As someone not simply committed to egalitarian justice, but living in an egalitarian community, Alf values a social life based on a shared understanding of justice with Betty. For this they must have common normative and empirical expectations about what, in their interactions, they take justice to require.[8] To say they have common "normative expectations" is to say that in a given interaction, Alf expects that Betty will hold that justice demands ϕ, and that Betty thinks Alf will hold that justice demands ϕ. Moreover, they have at least second-level common knowledge about this: Alf knows that Betty expects him to demand ϕ, and Betty knows the same about Alf. To say they have common "empirical expectations" is for them to expect the other to actually ϕ in that situation. Thus we are led to

> ASSUMPTION 3: THE COORDINATION COMMITMENT. As competent members of utopia, Alf values sharing normative and empirical expectations about justice with Betty, as she does with him.

The following analysis depends on an additional assumption, that is,

> ASSUMPTION 4: THE INEVITABILITY OF COMPETENT DISAGREEMENT. For any two competent members of utopia, Alf and Betty, when they employ "I conclude we ought reasoning," they will sometimes disagree about what justice requires in some circumstances *C*. Alf will hold that action ϕ is required by justice while Betty will hold that not-ϕ is.

Now some may wish, as they say, to get off the bus right here. It might be thought that the Justice Commitment (ASSUMPTION 2) simply precludes ASSUMPTION 4: those who reason purely on the basis of justice, purified of

[8] See Cristian Bicchieri, *The Grammar of Society* (Cambridge: Cambridge University Press, 2006), esp. chap. 1.

the corrupting effects of self-interest, will inevitably agree. At one point, for example, Cohen asks us to recall Hobbes's observation that in geometry

> men care not, in that subject, what be truth, as a thing that crosses no man's ambition, profit or lust. For I doubt not but if it had been a thing contrary to any man's right of dominion, or to the interest of men that have dominion, *that the three angles of a triangle should be equal to two angles of a square*, that doctrine should have been, if not disputed, yet by the burning of all books of geometry, suppressed, as far as he whom it concerned was able.[9]

So perhaps it is only self-interest that clouds our understanding of the results of moral geometry. Perhaps moral judgments clash only because our interests are engaged. If we reasoned without the corrupting effects of interests, perhaps private judgments about the requirements of justice would concur.

I do not wish to disparage the idea that, somehow, passions and interests are at the root of moral and political disagreement (which, of course, to some extent they surely are). This conviction is basic to Enlightenment thinking. It perhaps reaches its apex in the anarchistic utopianism of William Godwin: if individuals are educated to reason, they shall control (or overcome) their passions, and conflict will largely pass.[10] Under the influence of reason purified of passions, harmony of private judgment could be expected. More generally, many have supposed that the free inquiry of scientists and philosophers inevitably leads to agreement in private judgment because (i) the truth is the same for everyone, (ii) reason is a shared capacity of all human beings, and (iii) the norms of good reasoning are universal. Thus, people reasoning correctly about the world will arrive at the same answer. The true and valid results of one person's reasoning are thus necessarily true and valid for all. "Enlightenment philosophers," John Passmore noted, were convinced that "mankind had in the seventeenth century lit upon a method of discovery [the scientific method], a method which would guarantee future progress."[11] And as Isaiah Berlin observed, "in the eighteenth century there was a fairly wide consensus that what Newton had achieved in the region of physics could surely also be applied to the regions of ethics and politics."[12]

[9] Thomas Hobbes, *Leviathan*, ed. Edwin Curley (Indianapolis, IN: Hackett, 1994), 61 (chap. 11, para. 21). Cf. Cohen, *Rescuing Justice and Equality*, 6.

[10] See William Godwin, *An Enquiry Concerning Political Justice, and its Influence on General Virtue and Happiness,* two vols. (London: G. G. J. and J. Robinson, 1793). Thomas Malthus targeted Godwin's rather extreme claim that passions, including that for sex, would wither away. Thomas Robert Malthus, *Malthus — Population: The First Essay* (Ann Arbor, MI: Ann Arbor Paperback, 1959 [1798]).

[11] John Passmore, *The Perfectability of Man* (London: Duckworth, 1971), 200.

[12] Isaiah Berlin, *The Roots of Romanticism* (Princeton, NJ: Princeton University Press, 1997), 23.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

Yet despite its respected ancestry, the hope that real excellent moral reasoners might secure complete consensus in their judgments seems deeply implausible in light of the modern experience of pervasive disagreement. As Rawls tells us, such disagreement appears to be "the natural outcome of the activities of human reason under free enduring institutions."[13] A variety of analyses lead to this conclusion, such as Rawls's own doctrine of the burdens of judgment and Johnathan Haidt's account of the fundamentally diverse foundations of moral judgment.[14] I shall not rehearse them here. It is, though, worthy of note that moral disagreement seems more, not less, likely as moral reasoners become more sophisticated. A great deal of moral consensus in our society is the result of conventional moral thinkers, who tend to take their cues from the beliefs of their fellows — such "conformity bias" is in some ways the heart of human culture.[15] But the sort of sophisticated, autonomous, reasoners we would expect in utopia would almost surely be more akin to Lawrence Kohlberg's "stage six" reasoners, who think things through for themselves, and are far less tied to socially accepted norms.[16] To be sure, advanced stage-six reasoners may converge on some matters, such as the injustice of forms of discrimination, but the very sophistication and independence of their reasoning will lead them to make subtle, and inherently, controversial distinctions. Such reasoners are better at splitting hairs than splitting the difference. For example, they are apt to concur that gender discrimination is wrong, but not on whether affirmative action is a form of discrimination or, even more basically, what affirmative action and discrimination really are.

[13] Rawls, *Political Liberalism,* expanded edition (New York: Columbia University Press, 2005), xxiv.

[14] According to Rawls, reasonable judgments so often are at odds because: (i) the evidence is often conflicting and difficult to evaluate; (ii) even when we agree on the relevant considerations, we often weigh them differently; (iii) because our concepts are vague, we must rely on interpretations that are often controversial; (iv) the manner in which we evaluate evidence and rank considerations seems to some extent the function of our total life experiences, which of course differ; (v) because different sides of an issue rely on different types of normative considerations, it is often hard to assess their relative merits; (vi) in conflicts between values, there often seems to be no uniquely correct answer (Rawls, *Political Liberalism*, 56–57). For Haidt's view, see *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon, 2012).

[15] See Peter J. Richerson and Robert Boyd, *Not by Genes Alone* (Chicago: University of Chicago Press, 2005), chaps. 1–3. It has been said that children can be understood as "cultural sponges." Alex Mesoudi, *Cultural Evolution* (Chicago: University of Chicago Press, 2011), 15. We often do not understand precisely the benefits of our cultural practices, but because culture is largely transmitted via imitation, people often do not have to know why something is done, only that it is the done thing around here. Whereas intelligent primates such as chimps tend to figure out problems for themselves, human infants appear to have a much stronger tendency to simply copy what they observe being done, copying "stupid" acts which the chimp apparently sees as pointless. See Victoria Horner and Andrew Whiten, "Causal Knowledge and Imitation/Emulation in Chimpanzees (*Pan Troglodytes*) and Children (*Homo Sapiens*)," *Animal Cognition* 8 (2005): 164–81.

[16] See Lawrence Kohlberg, *The Philosophy of Moral Development* (New York: Harper and Row, 1981), part two.

Thus, while they may perhaps concur at the level of broad principles, they are unlikely to secure the more fine-grained agreement that coordination in actual interactions requires.[17]

Those who are confident that ASSUMPTION 4 is wrong seem committed to something along the lines of

> *The Moral-Epistemic Homogeneity of Utopia*. In utopia all competent moral agents would always come to the same judgments about the demands of justice in circumstances *C*. The reasoning of one person about justice always stands for the reasoning of all.

All competent reasoners endorse the exact same principles of justice, arrive at identical judgments as to how they are best interpreted, and concur on what specific actions they call for.

It is worth pointing out that the Moral-Epistemic Homogeneity of Utopia has a somewhat unsettling implication: if in utopia someone *does* dissent, *ipso facto* he is incompetent, no matter how thoughtful and careful his reasoning is. Suppose there is some non-negligible chance that a mutation — call him Winston — will appear in utopia's population, whose reasoning leads him to different conclusions about justice. Given the Moral-Epistemic Homogeneity of Utopia, we have conclusive evidence that Winston is not competent; he suffers from what might be called an ideological disability. It is not simply that he disagrees, but we have a conclusive inference *from* his disagreement *to* the conclusion that he suffers from "moral incompetency." He may be patiently told that competency is learning to think as others think, but suppose he resists. As he sees it, others are saying 2 + 2 = 5, and that simply cannot be right. Perhaps a humane Commonwealth of Bees would treat him with advanced psychiatric techniques — after all, it would not manifest communal concern to simply leave him to his disability. If his disorder could infect others, perhaps more drastic public health measures might be called for. The line between utopia and dystopia begins to look precarious.[18]

[17] I have considered in more depth the relation between agreement on principles and disagreement on interpretations, in my "On Justifying the Liberties of the Moderns," *Social Philosophy and Policy* 25 (2007): 84–119. It is important that in moral matters, when we disagree, we cannot suspend judgment; something must be done. It is not at all clear that such suspension is the proper policy even in purely epistemic matters. For a general overview of the "peer disagreement" literature in epistemology, see the essays in *The Epistemology of Disagreement*, ed. David Christensen and Jennifer Lackey (New York: Oxford University Press, 2013).

[18] A referee objects: if "you *already* believe Assumption 4, then you will rightly find it dystopian that dissenters would be treated so disrespectfully. But if you instead believe that *in utopia all competent moral agents would agree*, then you really are going to think the dissenter in utopia is incompetent. The dissenter should be regarded much the way we regard Holocaust deniers or flat earthers or rabid racists who advocate lynching — there's some serious failure of reasoning going on here and they really would be more competent if they were able to think as others think." The reader takes (for sake of analysis) the argument based on *The Moral-Epistemic Homogeneity of Utopia* as *modus ponens* — the conclusion indeed follows.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

C. *A toy analysis of the paradox*

Let us attribute to our egalitarians "preference orderings" in the formal sense that, given their wholehearted egalitarian commitments, they can rank alternative social states and actions.[19] These preference orderings reflect only our first three assumptions: a thorough commitment to acting on egalitarian justice as one understands it, and the value of coordinating with each other on the demands of egalitarian justice. I initially assume that the commitment to justice is fundamental: at least in utopia, a conscientious person would not be committed to acting contrary to what he or she understands as the demands of perfect justice. That, I take it, is an appeal of utopia: each excellent moral agent does what she believes is the exactly right thing to do. It is important to stress that self-interest does not enter into their preferences in any way whatsoever. I shall call the ranking of options that such preferences generate a *fully egalitarian utility function*.

We now invoke ASSUMPTION 4 on the inevitability of disagreement. Suppose, then, that Alf believes that egalitarian justice dictates action *a* rather than *b*, while Betty believes that it requires *b* rather than *a*. Figure 1 shows that, if rational, Alf and Betty each will go his or her own way.

In this game, each orders the outcomes: (1) we both do what (on my view) is just; (2) I do what is just and the other acts on the erroneous view; (3) I act on the erroneous view and the other acts on the correct view (at least someone does the right thing!); and (4) we both act on the erroneous view. In this game the sole equilibrium is that Alf acts on his view (*a*), and Betty acts on her view (*b*), of justice. At either of the coordination solutions (when both play *a* or both play *b*), one of the parties would do better by changing his or her move, and acting on his or her favored interpretation of justice. Thus, even though we have integrated the value of community (ASSUMPTION 3), they will fail to coordinate because each is first and foremost devoted to his or her "I conclude we ought" judgment about egalitarian justice. Each most values doing the perfectly right thing.

Interactions along the lines of Figure 1 are not always a worry. Suppose Alf and Betty are gourmets, and he thinks restaurant *a* is better than *b*, while she has the opposite ranking; suppose further that for each it would, as it were, be the icing on the cake to share the excellent meal with the other. But still, good food is good food, and that comes first, so they play

---

I am suggesting that the conclusion shows the argument form to be *modus tollens*: if any dissenting conclusion about justice implies that "there's some serious failure of reasoning going on here" we should reject the premises that got us to what I cannot help but see as an objectionably dystopian conclusion. This apparently confirms the adage that one person's *modus ponens* is another's *modus tollens*.

[19] To remind readers of a basic but fundamental point: preferences are simply pairwise rankings of alternatives, not a type of mental state (such as a desire) that generates such rankings; and utility is a mathematical representation of how well a choice satisfies an ordered ranking, not a sort of goal or value. See here my *On Philosophy, Politics, and Economics* (Belmont, CA: Wadsworth-Thomson, 2008), chap. 2.

		Betty: *a*	Betty: *b*
Alf	*a*	4th 1st	2nd 2nd
	b	3rd 3rd	1st 4th

Figure 1: Go your own way — two resolute egalitarian moralists

the game of Figure 1 and each goes his or her own way. Not a real worry. But with justice, or a communal ethos, what Alf ranks as best is that he and Betty act on view *a* of the ethos because that is what the ethos really calls for, while Betty ranks as best both she and Alf acting on view *b*, as demanded by the ethos. Their judgments are not "I conclude *I* ought" but "I conclude *we* ought." In an important sense, as Cohen himself stresses, justice is what we do together.[20] Still, if pushed to choose between acting alone on the correct view ("What egalitarian justice demands") and acting together on the inferior view ("What the other mistakenly thinks egalitarian justice demands"), each will choose to stay true to his or her egalitarian commitments, but this is hardly like each enjoying a fine meal on one's own.

The problem is severe. Alf's and Betty's moral systems are such that each most highly values egalitarian justice (as he or she understands it), and will always choose to act on correct egalitarian justice over flawed versions. But they so value egalitarian justice because they seek a way to live together — an egalitarian community — but if each acts on his or her own view of how they both should live under egalitarian justice, they cannot live together under common egalitarian justice. For Kant, this paradoxical situation of acting on justice but in so doing failing to really secure justice, characterizes the state of nature:

> Although experience teaches us that men live in violence and are prone to fight one another before the advent of external compulsive legislation, it is not experience that makes public lawful coercion necessary. The necessity of public lawful coercion does not rest on a fact, but on an a priori Idea of reason, for, *even if we imagine them to be ever so good natured and righteous* before a public lawful state of society is established, individual men, nations and states can never be certain they are secure against violence from one another because each will have the right to do what *seems just and good to him*, entirely independently of the opinion of others.[21]

[20] Cohen, *Rescuing Justice*, 175ff.

[21] Immanuel Kant, *The Metaphysical Elements of Justice*, 2nd edition, ed. and trans. John Ladd (Indianapolis, IN: Hackett, 1999), 116 [sect. 43]. Emphasis added.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

Kant goes on to insist that interpersonal justice is absent in the state of nature because each relies on his own judgment, and thus "when there is a controversy concerning rights (*jus controversum*), no competent judge can be found to render a decision having the force of law."[22]

Kant's great insight is that, even if one truly knows what justice is, and acts on it, justice is not really secured if no one else acts on one's view. Alf goes ahead and acts on correct justice, but Betty does not follow. In this case, a person may know justice, act on his conviction, and fail to secure just social relations. And in a community of committed, thoughtful egalitarians, each could justifiably think *she* is the one who has the correct view that others erroneously reject.[23]

D. Community for its own sake

Cohen remarks that under the egalitarian ethos each values cooperation with others for its own sake.[24] It is important that the Paradox of Pure Conscientiousness will not arise if people care sufficiently about cooperation for its own sake. Suppose Alf and Betty flip the relative importance of their Justice and Coordination Commitments: commonality of action is preferred to doing what one believes is perfectly just. In this case, it would be more important for each to act as the other did than to act on what he or she sees as the correct view of justice (thus violating ASSUMPTION 2). Rather than the unfortunate game of Figure 1, their interaction is along the lines of Figure 2.

Here, each ranks the options: (1) we both act on my favored conception; (2) we both act on your favored conception; (3) I on act on my favored conception and you act on yours; (4) I act on your favored conception and you act on mine. This is a version of the classic "Battle of the Sexes Game."[25] To return to our dining example, now Alf and Betty are more interested in a social evening than the best food: while each still would like both, in this case, when forced to choose they sacrifice fine food for good company. In terms of egalitarian justice, Alf and Betty each still rank as best acting on his or her own understanding of justice when the other also acts on it, but each ranks acting on the other's (supposed to be erroneous) understanding of justice above acting alone on his or her favored (supposed to be correct) conception. Note, though, we have overcome the Paradox of Pure Conscientiousness in a rather unsatisfying way, by supposing that when there is a conflict, people care more about acting on shared judgment (coordination) than following their conscientious judgment. If we

[22] Ibid.

[23] Of course each could seek to force the other to do the just thing, but then we have what Kant thought of as a lawless state of nature.

[24] Cohen, *Why Not Socialism?* 42.

[25] R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: John Wiley and Sons, 1957), 90–94.

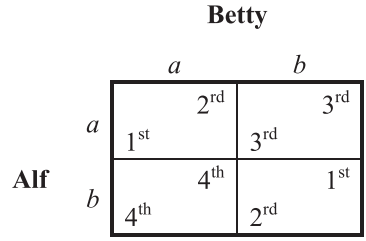


Figure 2: The battle of coordinating egalitarians

suppose that they are not, first and foremost, conscientious, and so when the crunch comes care more for coordination, it is not difficult to see how coordination can arise. If, however, there is any time when, rather than prioritizing coordination, they insist on acting in accordance with their own "I conclude we ought" judgments, they are back to Go Your Own Way.

E. Dropping non-accommodation (or adding cooperative justice)

It may seem that our difficulties stem from clause (ii) of the "I conclude we ought" View, which states that judgments of justice do not require accommodation to the deliberations of others. Let us, then, replace ASSUMPTION 1 with

> ASSUMPTION 1*: THE "I CONCLUDE WE BELIEVE THAT WE OUGHT" VIEW. As a competent member of utopia, if (i) Alf conscientiously deliberates and concludes that, given what he takes to be the correct view of justice and relevant empirical information, ϕ is required by justice under conditions *C*, where (ii) *this requires accommodating the views of others and meeting them half-way* as he understands them,[26] (iii) he reasonably concludes that justice instructs that we all ought to ϕ under conditions *C*, and so (*iv*) he justifiably ϕs in circumstances *C* and demands that others do so as well.

If Alf always follows these sorts of judgments, he still manifests conscientiousness, but his judgments build in the thought that no one can live with others simply on the terms he thinks most just. Now building in such accommodation almost certainly narrows the range of disagreement, and helps us move toward coordination. However, on its own, ASSUMPTION 1* fails to solve the Paradox of Pure Conscientiousness, for each person still acts on her view of "*I conclude* that we believe that we ought" and, once again, even the most excellent moral reasoners will disagree about what this is, for they can disagree as to what constitutes reasonable accommodation.

[26] This phrase comes from Rawls, *Political Liberalism*, 157, 163.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

A society in which all agree that we should make what, on each of our own views, is a reasonable compromise is by no means apt to uniformly concur about what that is. Indeed, some of the most intractable disputes can be about whether another person is appreciating our view and giving it due regard — think of how many marriages have run aground on those rocks. Thus an important insight: Figure 1's Go Your Own Way Game will often model our accommodating (ASSUMPTION 1*) moral reasoners. Simply suppose that Figure 1's orderings express Alf's and Betty's "I conclude that we believe that we ought" judgments. While it would certainly be wrong to say that this very significant modification as to what constitutes conscientious reasoning does nothing to mitigate our paradox, it certainly does not obviate it.

For much the same reason (perhaps, though, stronger)[27] it will not eliminate the Paradox to suppose that egalitarian justice includes the good of cooperation. Consider what we might call the "I conclude what a cooperative theory of justice says we ought" View. Granting that egalitarian justice requires cooperative social relations, our initial supposition was that in Alf's view Betty's advocacy of *b* is asking him to go against egalitarian justice by adopting unacceptably inegalitarian social relations: whatever role cooperation plays in his egalitarian conception of justice was already factored into the rankings in Figure 1. Generally, take any version of egalitarianism that requires the value of cooperation as part of egalitarian justice, and suppose Alf is deeply committed to *that*, and decides that it requires act *a*. Betty, equally committed to egalitarianism and its inclusion of cooperation, deliberates and judges *b* is called for. Figure 1, not Figure 2, still models important interactions.

F. The impossibility of a fully just conscientious community

The analyses based on ASSUMPTIONS 1 – 4, and 1* – 4 show that a community of excellent, conscientious, devotees of justice, who care first and foremost about acting justly but who also value moral community, must sometimes fail to secure perfect justice in their social relations. They will end up playing the Go Your Own Way Game, and this game characterizes less than fully just social relations. We thus have arrived at the impossibility of a fully just, fully conscientious utopian community.

III. PUBLIC JUSTICE

A. Social rules

That the accommodative account of moral reasoning fails to solve the Paradox of Pure Conscientiousness yields an important insight: the

[27] "Stronger" because this view does not build in special accommodation to the conclusions of others.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

Paradox cannot be overcome simply by expanding the content of "I believe we ought" type of reasoning to include how we ought to accommodate. Alf and Betty must share the same empirical and normative expectations about what the other will do, and what the other expects one to do. This commonality of expectations cannot be generated purely internally via their individual reasoning as depicted in the accommodative view (Section II.D). As I have argued elsewhere, for them to achieve justice in their social relations there must be some rule *R* such that Alf (i) expects Betty to act on *R*, and (ii) he believes justice calls for acting on *R*, and (iii) he believes that she holds that justice calls for acting on *R* (*mutatis mutandis*, for Betty).[28] Without (i), the empirical expectation, Alf will not achieve coordination, for he will not know what Betty is going to do; without (ii), his own normative endorsement, Alf will not hold action on this rule to be just; without (iii), the normative expectation, Alf will not think that Betty holds action on this rule to be just. I shall call a rule that meets these conditions for all members of society *S*, one of Public Justice in *S*.

Public Justice must be distinguished from simple agreement in a specific interaction on what both will take as just. A one-off agreement would not secure empirical expectations, and so each would not know what will be just next time. Under such conditions, it is doubtful that either could have well-grounded expectations as to whether there will continue to be just social relations: indeed, neither can be confident what the other will take to be justice in future interactions. Coordination would be highly imperfect, and thus they would fail to secure perfect justice. In principle Alf and Betty could solve this problem with a rule of Public Justice for a "society" *S* of only two, Alf and Betty, which would then determine a shared understanding of justice only in their dyadic interactions. But purely dyadic Public Justice is implausible on a variety of grounds: as communities grow, the number of dyads increases exponentially, and many relations are not dyadic, but involve interactions among large and varying numbers of people. For practical purposes, our concern is a rule of Public Justice shared by a reasonably large number of interacting individuals.

B. Is public justice consistent with pure conscientiousness? The special case of conventions

Is Public Justice consistent with purely conscientious judgment based on "I conclude we ought" reasoning? Under very restricted conditions, it is.

To investigate this, let us expand our model to a multiperson interaction, still restricted to wholehearted egalitarians, each of whom ranks acting on justice as he or she sees it as more important than coordinating with others, though that too is of value. (Again, only these two considerations

[28] See my *Order of Public Reason* (Cambridge: Cambridge University Press, 2011), 163–79. Compare Bicchieri, *The Grammar of Society*, chap. 1.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

enter into their utility function and, so, their decisions.) Rather than depicting the problem as one of dyadic interactions, we suppose a set *N* of wholehearted egalitarians is deliberating about rules by which to coordinate their actions. Now some rules are mere conventions that specify an empirical expectation about the behavior of others, such that given this expectation, each person's "I conclude we ought" judgment (acting only on one's private judgment of what is most just in that situation), can lead all to follow the same social rule. A wee bit more formally, let us say that *CR* is a conventional rule for society *S* if (i) those in *S* expect others to follow *CR* in the relevant circumstances; (ii) given this expectation, each person in *S* achieves her full (highest possible) egalitarian utility by acting on *CR*; and so (iii) no one in *S* has an incentive to defect.[29] Let us also suppose that each in *S* believes that others think that *CR* is the just rule or at least that no other rule is more just. On this characterization of a convention, if a group of people were setting out on an activity together with a division of labor apportioned among them by rule *CR*, and if each would maximize her own egalitarian utility by perfectly playing her part given her expectation that others would do so as well, *CR* would constitute a just conventional rule, allowing them to coordinate their egalitarian utility-maximizing actions and sharing just social relations. As Virgil recognized, the Commonwealth of Bees is characterized by a complex division of labor, but for all that, it is based solely on naturally harmonious individual judgments.[30] If, for example, each participated in a market scheme simply as a way to best satisfy his or her own fully egalitarian utility, we would still have a harmony of private judgment based on a shared egalitarian ethos.[31] So if the rules of Public Justice are *CR* rules, then purely conscientious moral agents can achieve just social relations. Each would follow his or her own "I believe we ought" judgment while being guided by a common conventional rule.

Cohen remarks that egalitarianism and its rules are characterized by indeterminacy.[32] Suppose, then, that our *N* wholehearted egalitarians

[29] This is a broad definition of a convention that does not require that Alf and Betty are playing a strict coordination game, though it encompasses such games. To get more precise, the maximization condition would have to be spelled out, but it would take us too far astray. On the narrower, more classic notion of a convention, see David Lewis, *Convention: A Philosophical Study* (Cambridge, MA: Harvard University Press, 1969). See also Robert Sugden, *The Economics of Rights, Co-operation and Welfare* (Oxford: Blackwell, 1986); Russell Hardin, *David Hume: Moral and Political Theorist* (New York: Oxford University Press, 2007), chap. 4.

[30] H. Musgrave Wilkins, *A Literal Translation of the Eclogues and Georgics of Virgil* (London: Longmans, Green and Co., 1871), 94ff.

[31] Sometimes this seems suggested by Joseph H. Carnes in his utopian egalitarian market scheme, as if the motivation by each to do her social duty would be enough. However, overall he sees the need for rather more explicit socialized incentives, a matter I take up in Section IV.C. See his *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory* (Chicago: University of Chicago Press, 1981), 8. The hope that a division of labor could be the spontaneous upshot of free egalitarian choices is an enduring theme in utopian thought, such as in Fourier's scheme or Bellamy's socialist utopia.

[32] Cohen, *Rescuing Justice*, 123.

confront three possible conventional rules {*a*, *b*, *c*} that would allow them to coordinate their pursuit of fully egalitarian utility, but they conclude that the principles of egalitarian justice about this matter are indeterminate: they simply cannot rank the rules in terms of justice. Assume, though, that they all accept that coordinating on any of the three rules is better than option *o*, going without a coordinating convention. Recall that we have specified fully egalitarian preference orderings as reflecting only (i) the value of acting on the correct view of egalitarian justice and (ii) the value of social coordination on justice, where (i) is prior to (ii). We suppose that in this case our egalitarians cannot rank alternatives on value (i), which means all they have left is the value of coordination. Our three alternative egalitarian rules thus constitute a maximal set for our *N* egalitarians in comparison with forgoing any rule about the matter at all: all agree that $a \succ o$,[33] $b \succ o$, and $c \succ o$. As Amartya Sen has shown, the rational thing to do is to choose one from this set, even though they cannot rank the elements within the set.[34] Note an interesting consequence: given that the most important egalitarian element of the utility function has been rendered irrelevant by the indeterminacy of their egalitarian commitments in relation to ranking the rules, their problem is now a coordination game, and they can seek a convention. Suppose that "1st" simply means "better than 2nd";[35] for ease of explication, consider the choice between just Alf and Betty as in Figure 3.

We see that Cohen is right: indeterminacy can indeed be the "saving grace" of radical utopian egalitarianism[36] — if it is thoroughgoing. Suppose, however, that while *N*-1 persons are unable to rank {*a*, *b*, *c*}, one person, Vladimir Ilyich, is fully confident that he can rank them, and does: $a \succ b \succ c$! Vladimir Ilyich now becomes decisive over all options in the original maximal set. Consider simply a two person interaction over two options, *a* and *b*.

Vladimir Ilyich's orderings are the resolute egalitarian moralist's from Figure 1, who holds that $a \succ b$. Betty is the indecisive egalitarian from Figure 3. Betty knows that Vladimir Ilyich will never play strategy *b*. It is, as game theorists say, strictly dominated by *a*: Vladimir Ilyich always does better by playing *a* than *b*. Betty knows this, and so knows the real game for her is only Vladimir Ilyich's top row, and if he plays *a* she should too. And he will, so *a* it will be. The remainder of *N* will reason in the same way as Betty. They have no reason to rank *a* above *b* or vice versa: they simply

[33] Read as "*a* is preferred to *o*."

[34] See, for example, Amartya Sen, "Maximization and the Act of Choice" in his *Rationality and Freedom* (Cambridge, MA: Harvard University Press, 2002): 159–205.

[35] Thus we cannot say that two first choices are "equal," that is, that the players are indifferent between, say, all play *a* and all play *b*. That would require a complete ranking. For those worried by the apparent assumption of indifference, and so completeness, in Figure 3, I consider at some length how to model this sort of choice under incompleteness in *The Order of Public Reason*, 303–10.

[36] Cohen, *Rescuing Justice*, 123.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

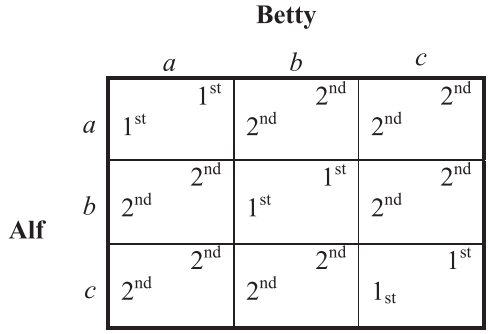


Figure 3: A coordination game resulting from indeterminacy

want to coordinate, and will do whatever allows them to best coordinate. Since everyone but Vladimir Ilyich will coordinate on either *a* or *b*, but Vladimir Ilyich will only coordinate on *a*, it maximizes everyone's benefits from coordination to follow Vladimir Ilyich's lead. Under conditions of perfect information, Vladimir Ilyich is an egalitarian dictator.

If there are *N*-2 resolute egalitarians then if they agree that *a*>*b*, they will together lead the others to *a* rather than *b*. We could think of them as forming a vanguard elite, jointly controlling the decision, and the above reasoning will hold. If they disagree, then coordination will begin to break down. When resolute Vladimir Ilyich meets Leon, the other resolute egalitarian with the opposite ranking, they will play the game of Figure 1, and each goes his own way. When the resolute egalitarians disagree, it is better for the indecisive egalitarians to follow the majority of resolute egalitarians; they will then maximize their coordination with others. So the resolute are still leading the way. However, as resolute egalitarians proliferate and they disagree among themselves, the indecisive egalitarians will confront increased instances of noncooperation. As the noncoordination interactions increase, the indecisive egalitarians' empirical expectations that others will act in the expected, conventional, way will also become weakened. As they witness increased instances of nonconventional behavior, the empirical expectations that are the heart of the convention, and so the convention itself, will be undermined.[37]

It may seem that this constitutes a counterexample to my impossibility argument: we have just shown that a society of conscientious moral agents can coordinate on justice, and thus the Paradox of Pure Conscientiousness apparently has been overcome. Note, however, that ASSUMPTION 4, inevitable disagreement, was not invoked. In Figure 3 there is no disagreement; the thoroughgoing incompleteness of judgment implies that no one has a ranking to base a disagreement on. In the case of Figure 4, with Vladimir Ilyich and Betty, we still do not have disagreement: Vladimir Ilyich has

[37] On the rule-destroying effects of undermining empirical expectations, see Cristina Bicchieri, *Norms in the Wild* (Oxford: Oxford University Press, 2016), chaps. 3–4.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

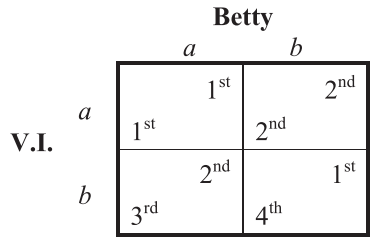


Figure 4: What is to be done? Go my way! — a resolute egalitarian makes the decision for all

a ranking but Betty does not. As soon as we confront an interaction that satisfies ASSUMPTION 4, with Vladimir Ilyich and Leon, we again see imperfect justice. This special convention-through-indeterminacy case thus has much in common with Figure 2, the Battle of Coordinating Egalitarians: because the moral judgment part of egalitarian utility function is essentially not relevant, we are back to a coordination game.

C. The Gap

The upshot of our analysis, then, is that purely conscientious agents who arrive at definite conclusions about the requirements of justice are unable to secure perfect justice. When our utopian moral agents arrive at conflicting judgments, they will be faced with what I shall call

> *The Gap*: When ASSUMPTIONS 1–4 (or 1*–4) are met regarding some matter of justice, and society *S* has a rule of Public Justice about this matter, some individuals will conclude that Public Justice falls short of what they view as perfectly just.

The Gap supposes that in a utopian community all can see the rule of Public Justice as in some sense a *bona fide* rule of egalitarian justice, and not simply an obvious injustice. This supposition was implicit in our coordination game in Figure 2: each saw the other's favored view as a rule of justice and not, for example, a brute expression of interest or ideology. We are, after all, talking about utopia in which everyone reasons excellently on the basis of justice. So I assume it makes perfect sense for Betty to say to Alf, "your *bona fide* judgment of justice falls short of perfect justice." This implies a Gap between what Betty deems perfectly just and what Alf does. The claim of The Gap is that, when ASSUMPTIONS 1–4 (or 1*–4) are met, and when the society has Public Justice on this matter, at least one person will see Public Justice as falling short of perfect justice as she sees it, and so she will have a Gap between her conscientious judgment and the rule of Public Justice. Once we posit Public Justice, the Paradox of Pure Conscientiousness is transformed into the inevitability of The Gap.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

But why should Betty, a conscientious moral agent, accept The Gap? It will not suffice to say that unless she accepts The Gap she and her fellows in the egalitarian community cannot share just social relations. As we saw in Section II.E, we can suppose that Betty has already factored the importance of this into her egalitarian utility; *ex hypothesi*, after doing all of this she confronts The Gap. So it looks as if when a purely conscientious agent confronts The Gap she must go her own way, but if she does so she will defect on Public Justice, and so she and her fellows will fail to secure justice in their social relations. Can we provide Betty with good moral reasons to accept The Gap?

IV. Public Justice and Incentives

A. Self-effacing reasoning and the artificial Commonwealth of Bees

Hobbes was perhaps the first to construct a social philosophy on the recognition that on *any* matter of judgment, when people seek to apply their reason they can come to different conclusions, and that in such situations coordination of judgment cannot be achieved by each insisting that her reason is *really, truly*, right reason.

> And, as in arithmetic, unpractised men must, and professors themselves may, often err, and cast up false, so also in any other subject of reasoning, the ablest, most attentive, and most practised men may deceive themselves, and infer false conclusions; not but that reason itself is always right reason, as well as arithmetic is a certain and infallible art, but no one man's reason, nor the reason of any one number of men, makes the certainty, no more than an account is therefore well cast up, because a great many men have unanimously approved it. And therefore, as when there is a controversy in an account, the parties must by their own accord, set up, for right reason, the reason of some arbitrator, or judge, to whose sentence they will both stand, or their controversy must either come to blows, or be undecided, for want of a right reason constituted by nature, so is it also in all debates of what kind soever.[38]

On the strongest interpretation of Hobbes's analysis, for Alf or Betty to have any rational view of justice at all, it must be endorsed by interpersonal rationality. David Gauthier maintains that "Hobbes's real concern in this passage . . . is surely with interpersonality . . . . Rationality frees us, not only from dependence on our passions,[39] but, *perhaps more remarkably, from dependence on our own considered judgments*, in contexts in which that

[38] Hobbes, *Leviathan*, 23 (chap. 5, para. 3).
[39] Recall here the passage from Hobbes in Section II.B, which Cohen quotes.

dependence is disadvantageous to us. In this respect, rationality is, as it were, the remedy for its own defects."[40] As Gauthier reads *Leviathan*, then, Hobbes maintains that it is irrational to insist that one's reasoning is equivalent to right reason in contexts of sustained disagreement.[41] Thus, Gauthier maintains, "on Hobbes's account, the individual mode of deliberation, in which each person judges for herself what she has reason to do, is *supplanted* by a collective mode, in which one person judges what we all have reason to do."[42] Hobbes "proposes an extension of reason, so that rationality comes to have a social dimension."[43]

This would solve Alf and Betty's conundrum in Figure 1 by abandoning ASSUMPTIONS 1–2. If there was some public process that both of their private judgments endorsed as the proper collective mode of rationality, then what this process deemed the requirements of true egalitarian morality would be accepted by each as the rationally correct requirements, and they would have little problem coordinating. They both follow the processes' output, and this defines Public Justice, giving the sole rational determination of what interpersonal justice requires. This process could invoke a variety of procedures. For Hobbes, it was the reason of the sovereign, but one could follow Rousseau, and see it as the reason of the general will:

> It is to law alone that men owe justice and liberty. It is this salutary organ of the will of all which establishes, in civil right, the natural equality between men. It is this celestial voice which *dictates to each citizen the precepts of public reason, and teaches him to act according to the rules of his own judgment, and not to behave inconsistently with himself*. It is with this voice alone that political rulers should speak when they command; for no sooner does one man, setting aside the law, claim to subject another to his private will, than he departs from the state of civil society, and confronts him face to face in the pure state of nature, in which obedience is prescribed solely by necessity.[44]

Thus, after a vote a minority would conclude that it has been shown wrong as to what the common good requires.[45] The minority accepts the

[40] Gauthier, "Public Reason," *Social Philosophy and Policy* 12 (1995): 19–42, at 27. Emphases added.

[41] Ibid.

[42] Ibid., 31. Emphasis in original.

[43] Ibid., 25.

[44] Jean-Jacques Rousseau, *A Discourse on Political Economy* in *The Social Contract and Discourses*, trans. G. D. H. Cole (London: J. M. Dent and Sons, 1923), 256–57. Emphasis added.

[45] "When therefore the opinion that is contrary to my own prevails, this proves neither more nor less than that I was mistaken, and that what I thought to be the general will was not so. If my particular opinion had carried the day I should have achieved the opposite of what was my will; and it is in that case that I should not have been free . . . . This presupposes, indeed, that all the qualities of the general will still reside in the majority . . . " (Rousseau, *The Social Contract*, in *The Social Contract and Discourses*, 112).

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

majority's judgment as its judgment, again ensuring convergence on the requirements of justice.

This solution requires thoroughly self-effacing personal reasoning when it comes to matters of justice in social relations. One reasons to the conclusion that one's reason must defer to reason as identified by public reason — so one must believe what the resulting Public Justice requires. This radically self-effacing account of private judgment ultimately seeks to create what we might call an artificial Commonwealth of Bees. Humans do not naturally possess identical private judgment about justice, but if our private judgments instruct us to accept and defer to a common public judgment, then for public purposes a very similar result could be secured. We think about public justice as one — following the law, the "celestial voice which dictates to each citizen the precepts of public reason."

Although Hobbes appears to suggest this as a possible solution to the conflict of private judgment, even he ultimately deems it impossible. In analyzing the special case of claims of prophecy, Hobbes insists that we cannot fully abandon our own reason: "we are not to renounce our senses and experience nor . . . our natural reason."[46] Humans cannot ultimately submit their "intellectual faculty to the opinion of any other man. . . . For sense, memory, understanding, reason, and opinion are not in our power to change, but always, and necessarily such, as the things we see, hear, and consider suggest unto us; and therefore are not effects of our will, but our will of them."[47] Belief is not voluntaristic: even if one wills to believe what others believe about justice, this cannot produce a belief about justice. Belief is not the effect of our will, but our will is the effect of belief, and our beliefs about justice differ. Not only is Hobbes right that human rationality is not so self-effacing, but even if we tried to endorse the proposal, in the end private judgment would remain firmly in control. For suppose everyone accepts the principle "Public Justice is definitive on matters of interpersonal justice." But — and this is a familiar point in Hobbes scholarship — whether this condition is met (that is, that Public Justice only applies in conditions *C* — to interpersonal questions) cannot itself be determined by Public Justice, for Public Justice only has competency once condition *C* is met. If I do not believe *C* is met, then I will deny that this is a matter for Public Justice.[48] Each must use her own private judgment to determine whether *C* is met and, so, whether Public Justice has come into play.

[46] Hobbes, *Leviathan*, 245–46 (chap. 32, para. 2).

[47] Ibid., 246. "But in any business, whereof a man has not infallible science to proceed by, to forsake his own natural judgment, and be guided by general sentences read in authors, and subject to many exceptions, is a sign of folly, and generally scorned by the name of pedantry" (ibid., 27 [chap. 5, para. 22]).

[48] I argue for this conclusion in some depth in my *Contemporary Theories of Liberalism: Public Reason as a Post-Enlightenment Project* (London: Sage, 2003), 71ff. See more generally Jean Hampton's *Hobbes and the Social Contract Tradition* (Cambridge: Cambridge University Press, 1986), chap. 7.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

But, if private reason determines when public reason is determinative, then disputes will arise when we apply our private judgment to *that* question. Thus, we disagree about whether, say, a decision to abort a fetus is a private matter to be decided by a woman's private judgment, or whether we require Public Justice to determine the just outcome. A version of the dispute in Figure 1 reappears.

B. A commitment to public justification: Incentives to depart from perfect justice

Cohen argues that in an egalitarian community individuals would hold themselves responsible for justifying their behavior to each other. "A justificatory community," he explains, "is a set of people among whom there prevails a norm (which need not always be satisfied) of comprehensive justification."[49] Now insofar as one is moved to engage only in behavior that can be so justified, one is essentially privileging public justification over fidelity to one's private judgment about perfect justice. In Figure 1 — the Go Your Own Way Game — Alf cannot justify his action to Betty; he can explain it and make it intelligible, but he cannot justify it to her as the just thing to do. If he privileges the justificatory norm over acting on his private judgment about justice, then the *ethos of justification* trumps the commitment to acting on his private judgment about justice. And then Public Justice would be effective.

This is certainly a plausible way to proceed, but its upshot is that a utopian community requires incentives for people to act against their "I believe we ought" judgments. As we have seen in our examination of the accommodation view (Section II.E), building accommodation to the reasons of others into one's "I believe we ought" judgments will not overcome the Paradox. We require Public Justice, which we suppose can be justified to others, and so gives them reasons to depart from acting on perfect justice as they understand it. This requires that people have incentives to overcome The Gap between Public Justice and their own "I believe we ought" judgments. And, as we have seen, whatever these incentives are, *they cannot come from one's view of perfect justice*, for it is pristine "I believe we ought" judgments that led to our problem. The incentives must come from outside one's commitment to perfect justice such that given these incentives, Betty deems acting on Public Justice to be superior to acting on perfect justice as she sees it. She overcomes The Gap. Betty considers the demand of Public Justice that she ϕs, and consulting her "I conclude we ought" reasoning," she disagrees that ϕ-ing is perfectly just. But, we say to her, conforming to Public Justice and ϕ-ing nevertheless is based on important values and commitments she reflectively endorses. These values and commitments might appeal to a broader valuing of just social relations and the practice of justice: to achieve real social relations that are

[49] Cohen, *Rescuing Justice*, 43.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

approximately just, she must forsake devotion to her own pure "I believe we ought" judgments, and seek to coordinate with others. For Cohen this helps solidify certain sorts of communal relations; on my account it provides the foundation for mutual accountability and fruitful cooperative relations (and community for those who value it). Thus, we are doing what strikes so many as objectionable: even in utopia we need to find incentives for Betty to depart from her deepest convictions about perfect justice, and accept the Gap between it and what public justice demands. In utopia almost everyone must forsake their judgments about perfect justice and so utopia must provide them with incentives to do so.

Moreover, the more our private judgments lead us to disagree, the wider the incentive system will have to be. Suppose we "incentivize" Betty to prioritize public justification by appealing to the value of community, but she judges that community is not very important in this context, or that the value of community would be better supported by acting on perfect justice; if so, we will fail to convince her to rank following Public Justice over acting on her "I believe we ought" judgment. However, if a wider "system of incentives" can be employed — appealing to the values of fruitful cooperation, security, and peace — it is more likely that Betty will have reason to concur that acting on Public Justice should be privileged — that she should accept The Gap.

Putting the matter in terms of a "system of incentives" perhaps sounds grubby and "economistic." But what we are doing in such cases is by no means grubby: we are seeking to provide reasons for individuals, who employ their reason, to arrive at different judgments of perfect justice, to endorse and comply with a public system of justice that (even in utopia) provides us with the great moral good of interpersonal relations based on mutually recognized requirements of justice. To think of this as grubby incentivizing (or akin to paying ransom) wrongly suggests that, if *they cared only about perfect justice* there would be no problem. But the problem would arise *precisely* if they did only care about such justice, in the sense described by ASSUMPTIONS 1–3.

C. *Moral incentives sans self-interest?*

It might be thought that, once again, I have simply defined the egalitarian ethos too narrowly. Let us return to the idea considered in Section II.D, that the ethos is not simply an overriding commitment to act on perfect egalitarian justice, but includes serving others as one is served by them. "Communal reciprocity" would characterize a utopian egalitarian community, and given such an ethos of service, each would freely defer to Public Justice as a way to best serve others.[50] We might understand this

[50] See Cohen, *Why Not Socialism?* 39ff. Or, as Sidney and Beatrice Webb put it, "The Worship of God is replaced by the Service of Man" (*Soviet Communism: A New Civilization*, third edition in one volume [London: Longman's, Green and Co., 1944], 913).

ethos of service to others or society as a "moral incentive." In his utopian egalitarian market, Joseph Carens uses "moral incentives as a major source of motivation for economic activities."[51] On a common view, moral incentives in economics are defined as "altruism + prestige."[52] Individuals perform out of an altruistic concern for others (or, perhaps, doing their social duty), and because they value the reputation achieved through being cooperative and socially concerned citizens. Thus, an egalitarian might say, while she can well admit the need to incentivize actions to comply with the requirements of Public Justice, what is critical is that this system of incentives builds on citizens' concern for their fellows and for their own social reputation, and so refuses to induce compliance by appeals to self-interest.

This, we might think, is still a uniquely socialist utopia. It is not a Commonwealth of Bees, but of humans, who require public rules and incentives, yet it clearly leaves behind appeal to self-interest. Interestingly, Carnes rejects this interpretation. It is, he says, "misleading to describe the moral incentives of the egalitarian order as altruistic. . . . People in the egalitarian system would have a different notion of self-interest from those in PPM [private property market] systems. Because of the differences in the socialization process they would value social approval more and income less."[53] To value one's social prestige and reputation as a good participant cannot be sharply distinguished from caring about self-interest.[54] It is more accurate to say that such systems seek to channel self-interest in different directions, where one's interest is intimately linked to the opinions of one's fellows and the services one provides to the community. Rousseau clearly saw this:

> every citizen shall feel the eyes of his fellow-countrymen upon him every moment of the day; that no man shall move upwards except by public approbation; that every post and employment shall be filled in accordance with the nation's wishes; and that everyone … should be so dependent upon public esteem as to be unable to do anything, acquire anything, or achieve anything, without it. The resulting emulation among all the citizens would awaken that patriotic fervor which raises men — as nothing else can raise them — above themselves.[55]

[51] Carnes, *Equality, Moral Incentives, and the Market,* 8.

[52] P. J. D. Wiles, *Economic Institutions Compared* (New York: Wiley, 1977), 27.

[53] Carnes, *Equality, Moral Incentives, and the Market,* 122.

[54] Cf. Cohen: "The Carnes system is Utopian partly because it relies entirely on non-self-interested choice" (*Why Not Socialism*? 65). This does not seem to be Carnes's view. More accurately, Carnes's system avoids appeal to what Wiles called "Benthamite incentives" — an exchange of work for some private good outside of working hours (*Economic Institutions Compared*, 15).

[55] Jean-Jacques Rousseau, *The Government of Poland*, trans. Willmoore Kendall (Indianapolis, IN: Bobbs-Merrill, 1972), 87. See also John W. Chapman, *Rousseau —Totalitarian or Liberal?* (New York: Columbia University Press, 1956), 60ff.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

Even moral incentives rely on self-interest, though it is not obvious that Rousseau's version of "moral incentives" — though probably entirely realistic — has much appeal. It is no bad thing, I think, to sometimes rely on plain old private self-interest as a foundation for endorsing, and complying with, Public Justice, even in an egalitarian utopia. Recall that the difficulty is that, if people are solely concerned with their "I believe we ought" judgments about perfect justice, they will not have sufficient incentive to comply with Public Justice when this entails a Gap. We must, then, expand the range of values that support the conclusions of Public Justice to overcome The Gap. However, so long as the only appeal is to normative concerns shorn of all self-interest, if our normative disagreements are wide, it will be doubtful whether good-willed individuals will wholeheartedly endorse Public Justice, and comply with its demands. Those focused only on normative matters may find themselves unable to overcome The Gap.

In important experiments, Harinck, De Dreu, and Van Vianen studied the contrast between (i) "*negotiation about interests*" that "arises when interdependent individuals or groups hold conflicting positions that are rooted in conflicting personal interests such as the attainment of money, time, personal benefits, or other scarce resources" and (ii) "*negotiation about an evaluative problem*" that "arises when interdependent individuals or groups hold incompatible positions that are rooted in different ideas about a problem that has no single demonstrably correct answer, such as a problem that involves norms and values."[56] In a study focusing on whether negotiating parties arrive at an agreement after thirty minutes, they found that although parties in interest-based disputes tend to see their aims as more directly opposed than those in evaluative-based disputes, negotiations about interest-based disputes were more likely to reach a consensual bargain. The key seems to be that people are willing to trade interests for the sake of an agreement, but are less willing to "trade" values and normative commitments, thus making it more difficult to reach a consensus on an acceptable outcome.[57] Insofar as Public Justice involves trading off what we care about for the sake of a mutually acceptable outcome (that is, overcoming The Gap), a dollop of interest can make accommodation more likely. Recall in this regard the fate of the utopian socialist community mentioned at the outset — New Harmony, Indiana. Founded in 1825,

[56] Fieke Harinck, Carsten K. W. De Dreu, and Annelies E. M. Van Vianen, "The Impact of Conflict Issues on Fixed-Pie Perceptions, Problem Solving, and Integrative Outcomes in Negotiation," *Organizational Behavior and Human Decision Processes* 81 (2000): 329–58, at 330. They also considered a third type of dispute, "a *negotiation about an intellective problem*," that is, where "interdependent individuals or groups hold incompatible positions that are rooted in different interpretations of an objectively verifiable problem" (ibid.). All emphases in original.

[57] See also Joshua Greene, *Moral Tribes: Emotion, Reason and the Gap between Them and Us* (New York: Penguin, 2013), 86–88.

like most Owenite communities it was characterized by internal disputes, effectively breaking into three communities in 1826, with the community effectively ending in 1827. A critical dispute was about the justice of inequality — one of the splinter groups was the "Community of Equality."[58] The important point here is that the appeal to self-interest is not required as a compromise with "grubby human nature" but as a way to induce high-minded people to accept Gaps so that all can share effective systems of public justice.

V. Why Not a Commonwealth of Bees?

We have all experienced interludes of synchronicity with others, where our private judgments perfectly match or complement each other. Like a perfect dance, we achieve lovely coordination, yet each acts solely on one's private judgment. This can occur when we most want to act as one, or have a good time together, whether it is a camping trip or dancing. Sometimes, alas, this joy of synchronicity takes on a less savory guise, as when people experience the joy of immersion into the Volk, or the solidarity of wartime.[59] What is impossible is to achieve a deep synchronicity in large groups through sustained harmony of judgments about perfect justice or morality. When we are having a good time, judgments about the best and the right take a back seat — our coordination interests trump our convictions about the best thing to do (Section II.D). As we saw, two gourmets are likely to eat alone (Section II.C), but two lovers are likely to eat together (probably not paying that much attention to the food). Immersion into the Volk is more ominous, for such synchronicity requires us to go places where even Hobbes feared to tread — total submission to the judgment of the Leader, abandoning our reason to follow his.

As Hobbes observes in our epigraph, seeing the beautiful, perfect, synchronicity of the beehive, many have asked why humans cannot achieve something like it — or at least aspire to it in their utopian dreams. I have argued that this is impossible for us. Not because we are not up to it, being stuck in the capitalistic mire "of greed and fear,"[60] but because when goodwilled humans apply their reason they arrive at different conclusions, even about justice. A utopian project that denies this ultimately must be one that rejects the motto of the Enlightenment — "Have courage to use

[58] See A. Haworth, "Planning and Philosophy: The Case of Owenism and the Owenite Communities," *Urban Studies* 13 (1976): 147–53, at 153; Krishan Kumar, "Utopian Thought and Communal Practice: Robert Owen and the Owenite Communities," *Theory and Society* 19 (1990): 1–35, at 18.

[59] This may be a more basic feature of humans than enjoying camping. See Greene, *Moral Tribes*.

[60] Cohen, *Why Not Socialism?* 40.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press

your reason!"[61] Sustained social synchronicity and the free use of one's own reason are ultimately incompatible. Any "ideal theory" that recognizes this basic truth must embrace incentivized Public Justice.

Philosophy, University of Arizona

[61] Immanuel Kant, "What is Enlightenment?" in *On History: Immanuel Kant*, ed. Lewis White Beck (Indianapolis, IN: Bobbs-Merrill, 1957): 3–10, at 3.

https://doi.org/10.1017/S0265052516000352 Published online by Cambridge University Press