Hume's theory of justice and Vanderschraaf's Vulnerablity Objection

Robert Sugden

University of East Anglia, Norwich NR8 6SL, UK

r.sugden@uea.ac.uk

8 November 2019

Peter Vanderschraaf's *Strategic Justice* is a major philosophical study of the concept of justice as mutual advantage, conducted within the framework of modern game theory (Vanderschraaf, 2018; henceforth *SJ*). Vanderschraaf sees himself as working in the tradition of David Hume's *Treatise of Human Nature* (1739-40/ 1978; henceforth *Treatise*) and *Enquiries concerning Human Understanding and concerning the Principles of Morals* (1748-51/ 1975; henceforth *Enquiries*); he describes Hume as among 'the greatest justice-as-mutual-advantage theorists' (*SJ*, xi–xii). I have worked in this tradition too, particularly as the author of *The Economics of Rights, Co-operation and Welfare* (Sugden, 2004; henceforth *ERCW*). Naturally, I am interested in similarities and differences between Vanderschraaf's analysis and my own.

I was particularly struck by Vanderschraaf's discussion of the 'Vulnerability Objection' to theories of justice as mutual advantage. Defining members of society as vulnerable if they are 'unable through their own efforts to contribute to the cooperative surplus', he takes it to be uncontroversial that if a putative theory of justice as mutual advantage 'denies the vulnerable any of its advantages', then that theory is 'no proper account of justice at all' (*SJ*: 281). Vanderschraaf argues that many theories of justice as mutual advantage, including Hume's, *do* deny the benefits of justice to the vulnerable. His preferred characterization of justice as mutual advantage allows (but does not require) the rules of justice to protect the vulnerable. His discussion of vulnerability focuses on a game that is created by adding vulnerable individuals to the 'Mutual Aid Game' that I introduce in *ERCW*.

I will argue that Vanderschraaf is overlooking crucial differences between what Hume intends his theory to do and what Vanderschraaf means by an 'account of justice'. I will try to show that the problem of vulnerability is less stark if one takes the more Humean approach of *ERCW*.

1. The problem of vulnerability

Modern theorists of justice as mutual advantage are often troubled by the thought that, if society were viewed only as a cooperative venture, people who were unable to contribute

anything to that venture would have no claim on the surplus that it produced.[1] A famous expression of that thought appears in Hume's *Enquiries*, and is important for Vanderschraaf's claim that the Vulnerability Objection applies to Hume's theory.

Before examining what Hume says, let me say that there are discontinuities between the *Treatise* and the *Enquiries*, and that I when I talk about Hume's theory of justice, I will mean the theory presented in the *Treatise*.[2] In the passage in the *Enquiries*, Hume is trying to show that 'public utility is the *sole* origin of justice' (*Enquiries*: 184). He presents a disturbing thought experiment:

> Were there a species of creatures intermingled with men, which, though rational, were possessed of such inferior strength, both of body and mind, that they were incapable of all resistance, and could never, upon the highest provocation, make us feel the effects of their resentment; the necessary consequence, I think, is that we should be bound by the laws of humanity to give gentle usage to these creatures, but should not, properly speaking, lie under any restraint of justice with regard to them, nor could they possess any right or property, exclusive of such arbitrary lords. (*Enquiries*: 190)

There are uncomfortable similarities between Hume's imaginary species and vulnerable groups in the real world. Vanderschraaf's category of the vulnerable includes infants, young children, the demented old and (a particularly troubling case) people who are permanently and severely disabled from birth (*SJ*: 280). Hume is clearly uncomfortable about the implications of his argument for (what was then) the relationship between native peoples and European settlers in the Americas. The Europeans, he says, have been

---

[1] Two examples: Having said that his account of justice is based on the idea that society is to be conceived as a fair system of cooperation, Rawls (2003: 20–21) says: 'Given this aim, I put aside for the time being these temporary disabilities [i.e. illness and accident] and also permanent disabilities or mental disorders so severe as to prevent people from being cooperative members of society in the usual sense' (p. 20) and 'In any case, we should not expect justice as fairness, or any account of justice, to cover all cases of right and wrong' (p. 21). Gauthier (1986: 268) acknowledges that moral constraints deriving from mutual advantage 'do not correspond in every respect to the "plain duties" of conventional morality. Animals, the unborn, the congenitally handicapped and defective, fall beyond the pale of a morality tied to mutuality.'

[2] Despite the fact that Hume 'cast anew' the arguments of the *Treatise* in his later *Enquiries*, asking that the latter be treated as the definitive statement of his philosophical principles (*Enquiries*: 2), I believe that work of the younger Hume is philosophically and psychologically deeper.

'tempted to imagine' that this relationship was like that between human beings and the imaginary species in his example. Giving in to this temptation, they have often 'thrown off all restraints of justice, and even of humanity' in their treatment of the native peoples (*Enquiries*: 191). Since Hume does not discuss this case any further, it is difficult to know how he would support the claim that the Europeans' actions were not only inhumane but also unjust.

Vanderschraaf interprets Hume as acknowledging that his theory of justice is inconsistent with the principle that 'vulnerable members of society are owed some benefits of justice' – a principle that Vanderschraff claims is a widely held 'considered judgment regarding justice' (*SJ*: 281, 283). I will ask whether there is a genuine inconsistency.

2. What is justice?

Vanderschraaf declares his objective on his first page:

> [H]ere is the thesis: *Justice is convention*. I believe that this justice-as-convention thesis, properly developed, provides the most cogent characterization of the general theory of justice as conceived as a system of rules for mutual benefit, or *justice as mutual advantage*. (*SJ*: xi)

But what does Vanderscrhaaf mean by 'justice' when he claims that justice *is* mutual advantage? He gives further clues when, still on the first page, he says:

> Plato set a template for all future philosophers by raising two interrelated questions: (1) What precisely is justice? (2) Why should one be just? One simple pair of answers to these questions had some currency even in Plato's Athens: (i) Justice is a system of requirements that are conventions, that is, mutually advantageous arrangements, of one's society, and (ii) Given their conventional nature, one has good reasons to obey these requirements simply because obedience serves one's own ends given others' expected obedience. (*SJ*: xi).

For Vanderschraaf, 'justice as mutual advantage' is an answer to questions (1) and (2); the content of that answer is (i) and (ii).

Plato's questions seem to presuppose that some moral entity called 'justice' already exists. Rather like European explorers who, after the coastline of Africa had been

mapped, set out to investigate the geography of the interior, philosophical explorers can investigate what the properties of justice really are. Vanderschraaf's discussion of the Vulnerability Objection implies that we can all express our 'considered judgements' about justice, knowing what it is that we are expressing judgements about.

This way of thinking about justice is foreign to the logic of Hume's *Treatise*. Hume presents the *Treatise* as a work of experimental psychology. Its full title is: *A Treatise of Human Nature: being an attempt to introduce the experimental method of reasoning into moral subjects*. According to the Introduction, it is to be a contribution to the 'science of man'– a subject that is fundamental to the all the other sciences, among which Hume includes philosophy. Its methodology is to be that of the natural sciences, based on 'experience and observation' (*Treatise*: xvi). Hume's discussion of justice is in Book III of the *Treatise*, entitled 'Of Morals'. Here, the objective is to achieve an empirical explanation of moral feelings. Hume famously argues that vice and virtue are not matters of fact, referring to objects outside our minds; the only matter of fact about vice is the 'feeling or sentiment of blame from the contemplation of it' (*Treatise*: 468–9). That is all there is to moral obligation:

> All morality depends upon our sentiments; and when any action, or quality of the mind, pleases us after a certain manner, we say it is virtuous; and when the neglect, or non-performance of it, displeases us after a like manner, we say that we lie under an obligation to perform it. (*Treatise*: 517)

Hume wants to explain the sense of moral obligation in terms of more basic human emotions of sympathy. He thinks that such an explanation is unproblematic for the 'natural virtues' of benevolence, humanity and parental care:

> Tho' there was no obligation to relieve the miserable, our humanity wou'd lead us to it; and when we omit that duty, the immorality of the omission arises from its being a proof, that we want the natural sentiments of humanity. A father knows it to be his duty to take care of his children: But he has also a natural inclination to it. (*Treatise*: 518–519)

The example of parental care is particularly significant, because Vanderschraaf explicitly includes young children in the category of the vulnerable. It seems safe to assume that, in Hume's view, the obligation of the father to take care of his children – that is, the sense of blame that attaches to a man who wilfully fails to meet this obligation – is just as

strict as, say, the obligation in justice to repay a debt. Hume is serious too about the obligation to relieve destitution: 'A rich man lies under a moral obligation to communicate to those in necessity a share of his superfluities' (*Treatise*: 482). We should not assume that Hume is thinking of this merely as an imperfect duty of charity. In eighteenth-century Britain, every parish was required by law to provide relief to destitute inhabitants, and to provide care for 'pauper children'; these activities were financed by local property taxes. Hume may have seen these practices as codifications of the obligation of the relatively well-off to contribute to the support of the desperately poor. Nevertheless, for Hume these are 'natural obligations', not obligations *of justice*. The difference is not a difference of moral force or of legal status. It is a difference between the causes of the feelings that constitute the moral obligations.

I maintain that the Platonic reading of the question 'What precisely is justice?' is out of place in a discussion of Hume's theory. If the Vulnerability Objection is rephrased as claiming that Hume's theory denies that the vulnerable are owed any moral obligations, one immediate answer is that there are obligations of humanity as well as of justice.

3. Hume's two-stage theory: the origin of justice

Hume has a special interest in obligations of justice because they present a challenge for his sentiment-based theory of morals. For Hume, 'justice' is simply the customary name for a particular subset of the set of general rules that are in fact recognized as morally obligatory in most societies. The rules of justice are those rules that secure three fundamental social practices: 'stability of possession', 'transference of property by consent', and 'performance of promises' (*Treatise*: 484–534). Significantly, Hume refers to these rules as '*what we now call* the rules of justice and equity' (*Treatise*: 505, emphasis added).

The challenge is that, although these rules are in fact perceived as moral obligations, it is implausible to assume that human beings have a natural inclination to conform to them. Actions that are *naturally* perceived as virtuous directly induce positive emotions, such as the pleasure of sympathy. In contrast, the actions that are

required by a rule of justice are virtuous only indirectly: it is only by recognizing the value of the rule that we can recognize the value of individual instances of conformity to it (*Treatise*: 477–484). Thus, an explanation of the obligation of justice must be a two-stage theory. It must first explain how there can be general conformity to the rules of justice, prior to those rules being perceived as morally obligatory. It must then explain why, as Hume puts it, 'we annex the idea of virtue to justice, and of vice to injustice' (*Treatise*: 498).

The first stage of Hume's theory works by way of his analysis of *convention*, which he presents in relation to the rule of stability of possession (*Treatise*: 484–501). The core features of this analysis can be reconstructed by using an idealized game-theoretic framework developed by David Lewis (*Convention*: 36–42). This framework is simpler than the one used by Vanderschraaf, but it is more transparent and is sufficient for my purposes.

Suppose that the members of some population recurrently engage with one another in some given type of interaction. (A single episode of interaction need not involve all members of the population, but every individual has some chance of interacting with every other.) This interaction (the *recurrent game*) is represented as a game in strategic form. Following Lewis, I treat payoffs as 'rough indications of strength of preference'. I make no distinction between preference and Hume's concept of 'interest', and do not make it a matter of definition that individuals always act according to their preferences. In Lewis's terminology, any list of strategies, one strategy for each player role in the recurrent game, is a (potential) *regularity* of behaviour in the population.

Now consider any regularity *R* such that the following three conditions are true and, in the sense defined by Lewis (1969: 52–60), common knowledge: [3]

(C1) Everyone conforms to *R*;

(C2) Everyone expects everyone else to conform to *R*;

[3] Lewis's concept of 'common knowledge' is formulated in terms of what people *have reason to believe*, not what they actually believe or know. On this, see Cubitt and Sugden (2003).

(C3) Everyone prefers to conform to *R*, conditional on everyone else conforming.

Conditions C2 and C3 are equivalent to saying that *R* is a Nash equilibrium in the recurrent game. Translated into Lewis's framework, Hume's description of the rule of stability of possession clearly satisfies C1, C2 and C3.

However, Hume wants to say more than this. He is not using 'convention' merely as a technical term. In its linguistically proper sense, 'convention' implies agreement. Hume wants to say that the rules that he calls conventions are *like agreements*. The rule of stability of possession, as he has described it,

> … may properly enough be call'd a convention or agreement betwixt us, tho' without the interposition of a promise; since the actions of each of us have a reference to those of the other, and are perform'd upon the supposition, that something is to be perform'd on the other part. (*Treatise*: 490)

The idea of agreement is important for Hume's suggestion that this rule can emerge spontaneously in a society in which self-interested individuals interact recurrently:

> Nor is the rule concerning the stability of possession the less deriv'd from human conventions, that it arises gradually, and by a slow progression, and by our repeated experiences of the inconveniences of transgressing it. (*Treatise*: 490)

Hume does not explain this process, but one might imagine it originating in, and gradually expanding from, implicit agreements among small groups of self-interested individuals who find it mutually advantageous to respect one another's possessions.

The additional property that establishes *R* as a convention is a condition of mutual advantage. In the case of justice:

> ['T]is certain, that the whole plan or scheme is highly conducive, or indeed absolutely requisite, both to the support of society, and the well-being of every individual. ... And even every individual person must find himself a gainer, on ballancing the account; since, without justice, society must immediately dissolve...' (*Treatise*: 497).

The implication is that benefit is to be measured relative to the state of affairs that would exist if the relevant interaction were not governed by *any* rule. Let us call this the *no-rule*

baseline. If we bracket out the problem of specifying exactly what is meant by 'not being governed by any rule', we can state a fourth condition simply as:

> (C4) The state in which everyone conforms to *R* is preferred by everyone to the no-rule baseline.

A *Humean convention* can be defined as a regularity *R* such that there is common knowledge that C1, C2, C3 and C4 hold.

Hume argues that each of the three rules of justice satisfies these conditions. In doing so, he claims to have fully explained 'the *natural* obligation to justice, *viz.* interest' (*Treatise*: 498). In other words, he has explained how a state of affairs in which the members of a society conform to the rules of justice could come into existence and persist over time. This explanation has not relied on any assumptions about morality; individuals have been assumed to act only on self-interest.

Stripped to its most basic features, and applied to a society of non-vulnerable people, Vanderschraaf's characterization of justice as convention is the conjunction of C1, C2, C3 and a particular version of C4.[4] The distinctive feature of Vanderschraaf's C4 is its definition of the no-rule baseline. This definition is complicated and its rationale is not fully explained, but the intuitive idea is that the baseline is a state of 'free-for-all' interaction in which each individual tries to gain at the expense of the others, without *any* restraint – not even of prudence, caution or rationality (*SJ*: 112). Whether or not this is what Hume has in mind, Vanderschraaf's chararcterization of justice is broadly similar to Hume's analysis of the natural obligation to justice. But Vanderschraaf does not address the question that Hume's analysis is intended to answer: Why, as a matter of empirical psychology, is justice perceived as obligatory? That is the second stage of Hume's theory.

4. Hume's two-stage theory: annexing the idea of virtue to justice

[4] See *SJ*, pp. 273–280. Vanderschraaf states an additional condition which (roughly speaking) requires that each person exercises *some* restraint. Given the nature of the free-for-all baseline, this is not very demanding. In the final few pages of *SJ*, Vanderschraaf proposes a further condition to eliminate equilibria that are not 'genuinely just'. However, he offers little explanation of why, if justice *is* mutual advantage, this condition is one of genuine justice.

At this point in his argument, Hume switches from an analysis of the *origin* of justice in small societies to an analysis of the *maintenance* of justice in larger ones. He continues to assume common knowledge that C1, C2, C3 and C4 hold (with 'preference' interpreted as interest, and perhaps with 'almost everyone' substituted for 'everyone'). However, he does not claim that conformity to the rules of justice can be sustained entirely by self-interest. That is true 'on the first formation of society', before society has 'become numerous, and has encreas'd to a tribe or nation'. But in a large society, each person's interest in the general practice of justice is more remote. In consequence, each individual is tempted to commit unilateral violations of the rules of justice. Perhaps surprisingly, Hume still claims that such violations are contrary to the individual's long-term interest, because of their tendency to undermine the general practice of justice.[5] But because people can be 'blinded by passion, or byass'd by [some] contrary temptation', C3 is not sufficient to motivate conformity to the rules of justice (*Treatise*: 498–501).

Nevertheless, by virtue of C4, unilateral violations of justice are still 'prejudicial to human society', and so evoke sentiments of disapproval from unbiased bystanders. These sentiments are induced by natural psychological mechanisms: a bystander who observes one person *i* in danger of being treated unjustly by another person *j* 'partakes of [*i*'s] uneasiness by *sympathy*'. In Hume's sentiment-based theory, this amounts to saying that everyone is under an obligation to conform to the rules of justice. Hume does not claim that this sense of obligation, when combined with interest, is sufficient to *motivate* conformity: he argues that large societies need governments to enforce the rules of justice (*Treatise*: 534–539). Nevertheless, the rules that governments are enforcing are moral obligations. Hume concludes: '*Thus self-interest is the original motive to the establishment of justice: but a* sympathy *with public interest is the source of the* moral approbation*, which attends that virtue*' (*Treatise*: 498–500).

[5] This is surprising, given that Hume recognizes the free-rider problem in his analysis of voluntary contributions to the public good of draining a meadow (*Treatise*: 538). Hume may have had second thoughts about his claim that, in large societies, unilateral violations of justice are always contrary to self-interest. In the *Enquiries*, he does not contradict the 'sensible knave' who thinks that an act of injustice 'will make a considerable addition to his fortune without causing any considerable breach in the social union and confederacy' (*Enquiries*: 282).

5. Normative expectations

Lewis's *Convention* (1969) was the first game-theoretic reconstruction of Hume's theory of justice. In *ERCW*, I build on Lewis's work, combining it with ideas from evolutionary game theory. Lewis offers an analysis of convention, and says that this 'turns out to be' the concept that Hume uses in his discussion of the origin of justice and property (*Convention*: 3–4). His definition of convention uses C1, C2 and C3, but replaces C4 with:[6]

> (C5) Conditional on his own conformity to *R*, each individual prefers (i) the state in which every other individual conforms to *R* to (ii) any state in which exactly one other individual fails to conform to *R*.

A regularity that satisfies C1, C2, C3 and C5 is a *Lewisian convention*.

Hume's C4 is based on a comparison between universal conformity to *R* and *universal non-conformity*. In contrast, C5 compares universal conformity with *unilateral non-conformity*. For example, think of the British regularity 'Drive on the left-hand side of the road'. To say that this satisfies C4 is to say that everyone prefers that this rule exists, rather than that there is no rule to keep vehicles from collision courses. To say that the rule satisfies C5 is to say that every driver who keeps left prefers that no other individual driver fails to keeps left. C4 assesses the rule *from outside* – from a viewpoint that considers the possibility that it might not exist. C5 assesses the rule *from inside* – from a viewpoint that assumes its existence. Of these two conditions, C4 is more suitable for an explanation of how conventions originate as tacit agreements in small societies, but C5 is more suitable for an explanation of why, in large societies, conformity to ongoing conventions is perceived as morally obligatory.

Lewis uses C5 to argue that conventions are 'a species of norms: regularities to which we believe one ought to conform' (*Convention*: 97–100). In *ERCW* I defend a generalization of this argument: if any regularity *R* satisfies C1, C2 and C5, it creates a

---

[6] See Lewis's 'first, rough, definition' of convention (*Convention*: 42). Lewis imposes the additional condition that there is some (potential but not actual) regularity *R′*, distinct from *R*, that also satisfies C3 and C5. This condition expresses an aspect of the ordinary-language meaning of 'convention', but it plays no role in Lewis's analysis of normativity.

normative expectation of conformity – the perception that conformity to *R* is a moral obligation. The underlying psychology is the natural feeling of resentment that is cued when one person's empirically well-grounded expectation about another person's behaviour is frustrated by the latter's deliberate and harmful action, combined with the sympathy that *R*-following bystanders feel for that resentment (Sugden, 1998; *ERCW*: 149–165, 214–218).[7]

6. The Mutual Aid Game

In *ERCW* I describe and analyse a Mutual Aid Game whose structure generalizes Hume's famous example of the two farmers who need each other's help in bringing in their harvests (*Treatise*: 520–521). My game is presented as a model of a practice followed in some working-class communities in Britain before the development of the Welfare State. If one member of a group of co-workers was unable to work because of sickness or injury, he was supported by money raised from the rest of the group in a voluntary collection. In my model, there is a group of workers who interact in each of a potentially infinite series of periods, but with some constant probability that the game will end after each period. In each period, one randomly selected worker is sick, and each other worker chooses whether or not to contribute to a collection. The utility cost of each contribution is significantly less than the utility benefit it gives to the recipient. Consider the following regularity *R*. At the start of the game, each group member is deemed to be in 'good standing'. In each period, if and only if the sick worker is in good standing, all other workers contribute. A worker loses his good standing if he fails to contribute in any period in which *R* requires this, but regains it after any period in which he has made a required contribution. If the probability that the game ends is not too high, *R* is a Lewisian convention (*ERCW*: 127–132). (On a natural interpretation of the no-rule baseline as the absence of the practice, *R* satisfies C4, and so is a Humean convention too.) Thus, if everyone can in fact be expected to conform to *R*, conformity is also a normative expectation. In Humean terms, it is a moral obligation.

---

[7] The analysis of resentment in *ERCW* draws on Adam Smith's (1759/ 1976) analysis of fellow-feeling. The idea that people are motivated to confirm other people's expectations appears in some later theories of 'social preferences' (e.g. Pelligra, 2005; Battigalli and Dufwenberg, 2007).

The essential idea of Vanderschraaf's 'Provider-Recipient Game' can be represented by revising the Mutual Aid Game so that a small proportion of group members are *vulnerable*: they have the same liability to sickness as the others, but are permanently unable to make contributions (*SJ*: 287–292). Consider the regularity R' that differs from R in only one respect: vulnerable individuals are always deemed to have good standing. If the proportion of vulnerable individuals is not too large, R' is a Lewisian (and Humean) convention. Thus, if everyone can be expected to conform to R', conformity to R' – and thereby, giving aid to the vulnerable – is a moral obligation in Hume's sense.

Vanderschraaf interprets this result as showing that obligations to the vulnerable can be *part of* (and not merely *coexist with*) a system of justice as mutual advantage. I am not persuaded. It seems self-evident to me that, in the revised Mutual Aid Game, the relationship between the vulnerable and the non-vulnerable is *not* mutually advantageous. One way of seeing why not is to suppose that practices of mutual aid originate in, and expand out from, tacit agreements between pairs of self-interested individuals (such as Hume's two farmers). It is clear that there would be no such agreement between a person who was capable of giving aid and one who was not.

Nevertheless, it is easy to imagine how, once a practice of mutual aid between non-vulnerable individuals has become established – or perhaps even once the possibility of such a practice has been recognized – natural psychological processes could lead to the inclusion of the vulnerable. One possibility is suggested by Hume's idea (offered as an explanation of why people tend to disapprove of their own breaches of obligations) that '[t]he *general rule* reaches beyond those instances, from which it arose' (*Treatise*: 499). In the case of mutual aid, it is a small mental step to reinterpret the rule 'Everyone who is not sick should contribute' as 'Everyone who is not sick should contribute *if they can*'. Perhaps unintentionally, Vanderschraaf suggests a different possibility when he says that it is a widely accepted considered judgement that 'the vulnerable are not to be denied *the benefits of* justice merely because they are vulnerable' (*SJ*: 281, emphasis added). Notice that this formulation (unlike some others that Vanderschraaf uses) does not assert that the vulnerable are entitled to those benefits *as a matter of justice.* It might more naturally be

interpreted as asserting an obligation of humanity. Non-vulnerable individuals have set up a scheme that provides each of them with a vitally important form of insurance. It would be inhumane for them to deny the benefits of this scheme to those of their fellows who, through no fault of their own, are unable to meet its entry requirements.

By adopting the revised good-standing rule, everyone meets an obligation of humanity. Once that rule is in place, it is in everyone's self-interest to conform to it, and everyone's conformity induces in everyone a sense that conformity is morally obligatory. What more need one ask? It seems that Vanderschraaf wants to be assured that the rule is a rule of justice. As a disciple of Hume, I am not sure what that assurance would mean, nor why it would matter.

References

Battigalli, Pierpaolo and Martin Dufwenberg (2007). Guilt in games. *American Economic Review: Papers and Proceedings* 97: 171–176

Cubitt, Robin and Robert Sugden (2003). Common knowledge, salience and convention: a reconstruction of David Lewis's game theory. *Economics and Philosophy* 19: 175–210.

Gauthier, David (1986). *Morals by Agreement*. Oxford: Oxford University Press.

Hume, David (1739-40/ 1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hume, David (1748-51/ 1975). *Enquiries concerning Human Understanding and concerning the Principles of Morals*. Oxford: Oxford University Press.

Lewis, David (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Pelligra, Vittorio (2005). Under trusting eyes: the responsive nature of trust. In Benedetto Gui and Robert Sugden (eds), *Economics and Social Interaction*. Cambridge: Cambridge University Press, pp. 195–124.

Rawls, John (1993). *Political Liberalism*. New York: Columbia University Press.

Smith, Adam (1759/ 1976). *The Theory of Moral Sentiments*. Oxford: Oxford University Press.

Sugden, Robert (1998). Normative expectations: the simultaneous evolution of institutions and norms. In Avner Ben-Ner and Louis Putterman (eds), *Economics, Values, and Organization*, pp. 73–100. Cambridge: Cambridge University Press.

Sugden, Robert (2004). *The Economics of Rights, Cooperation and Welfare*, second edition. Basingstoke: Palgrave Macmillan. (First edition 1986.)