

# 9 HUME'S NATURAL HISTORY OF JUSTICE

Mark Collier

## The Puzzle of Cooperation

Nature appears to have exercised, according to Hume, particular 'cruelty' towards human beings (T 3.2.2.2/484). When one surveys the rest of the animal kingdom, one finds a harmonious balance between *what creatures want* and *what they are able to do*. Lions have voracious appetites, but they have the power to satisfy them; sheep have simple desires, but these are easily fulfilled. It is in man alone that we discover, as Hume puts it, an 'unnatural conjunction of infirmity, and of necessity' (T 3.2.2.2/485). Upon further reflection, however, we can see that nature has provided us with a remedy for this unfortunate predicament: social cooperation (T 3.2.2.3/485).

'Tis by society alone he is able to supply his defects, and raise himself up to an equality with his fellow-creatures...By the conjunction of forces, our power is augmented: By the partition of employments, our ability encreases: And by mutual succour we are less expos'd to fortune and accidents. (T 3.2.2.3/485)

Social cooperation allows us to compensate for our feeble frames. We are not the fastest, strongest, or sturdiest creatures, but we make up for these shortcomings by joining forces with one another to an extent unmatched in the rest of the animal kingdom.

This raises an important question: What is it about human beings that allows us to cooperate on such a vast scale? Hume recognizes that the origin of human cooperation is a puzzle. Given the circumstances in which our primitive ancestors must have found themselves, one would expect interpersonal conflict to have been widespread. Resources in this ancestral environment were scarce, after all, and could not satisfy the desires of everyone. The agents who pursued these goods, moreover, would have been driven by unbridled appropriative impulses.

This avidity...of acquiring goods and possessions for ourselves and our nearest friends, is insatiable, perpetual, universal, and directly destructive of society. There scarce is any one, who is not actuated by it; and there is no one, who has not reason to fear

from it, when it acts without any restraint, and gives way to its first and most natural movements. (T 3.2.2.12/491–2)

When creatures with unrestrained appetites compete for limited resources, it seems likely that they would inevitably come to blows. Thus, it is *prima facie* difficult to understand how our ancestors cooperated with one another in these circumstances.

### Property Conventions

Hume maintains that one cannot simply appeal to generosity or benevolence in order to solve this puzzle. He does not deny that we are capable of acting unselfishly; the problem is that human kindness is narrowly tailored (T 3.2.2.5/487). We care about the pain and suffering of our friends and relatives, in other words, but we are also largely indifferent to the plight of those outside our close circle (T 3.2.2.6/487). It is easy to understand why we refrain from the possessions of our loved ones, then, but it is hard to see why we do so with regard to strangers, whose welfare is of little concern to us.

The key to solving this puzzle, according to Hume, lies with our capacity for strategic rationality. Our ancestors began to cooperate with strangers, he proposes, because they recognized that doing so would promote their mutual interests.

Men being naturally selfish, or endow'd only with a confin'd generosity, they are not easily induc'd to perform any action for the interest of strangers, except with a view to some reciprocal advantage, which they had no hope of obtaining but by such a performance. (T 3.2.5.8/519)

Hume maintains that the benefits of social cooperation are easy to discern. Our ancestors had 'repeated experience' with the 'inconveniences' of an uninhibited pursuit of scarce resources (T 3.2.2.10/490). These costs stood in stark contrast, moreover, with the benefits of cooperation displayed within the family (T 3.2.2.4/486). It would have required only minimal intelligence to recognize, therefore, the advantages of mutual restraint from the possessions of others. The benefits of this strategy, as Hume puts it, are 'palpable and evident, even to the most rude and uncultivated of human race' (T 3.2.7.1/534).

It is not enough to believe, of course, that there is reciprocal advantage in abstaining from the possessions of others. There is no use in adopting such a strategy unless others are willing to do the same. Thus, one must also believe that *others believe* that they stand to benefit from a policy of mutual restraint.<sup>1</sup>

This convention...is only a general sense of common interest; which sense all the members of the society express to one another, and which induces them to regulate their conduct by certain rules. I observe, that it will be for my interest to leave another

in the possession of his goods, *provided* he will act in the same manner with regard to me. He is sensible of a like interest in the regulation of his conduct. When this common sense of interest is mutually express'd, and is known to both, it produces a suitable resolution and behavior. (T 3.2.2.10/490)

Our ancestors began to act on their acknowledged interests, on this account, because they expected that others would do so as well. Property conventions were gradually established, according to Hume, as these mutual expectations gave rise to regular patterns of behaviour.

Humean conventions involve situations where 'the actions of each of us have a reference to those of the other' (T 3.2.2.10/490). Let us call individuals 'strategically rational agents' if they have the capacity to understand that the outcomes of their decisions often depend on the choices of others. We can say that agents enter into *conventions for the stable possession of property* if and only if:

- (A) Each agent believes that the other is strategically rational.
- (B) Each agent believes that mutual abstinence from the possessions of others would promote their individual interests.
- (C) Each agent believes that others believe that this policy would promote their individual interests.
- (D) The beliefs in (A)–(C) produce a suitable regularity in their behaviour.

This analysis makes it clear that property conventions require a fair amount of cognitive sophistication. Agents must not only make complex calculations about their own interests, but they must also entertain higher-order beliefs about the beliefs of others.

The only question that remains, according to Hume, is how our ancestors established particular rules for delineating possessions. How could strategically rational agents select among alternative methods of conferring ownership? It is at this point that Hume appeals to the considerations of psychological salience. Resources would have been assigned according to the rule of *present possession*, for example, since custom and habit naturally lead us to value goods with which we are acquainted more highly than those enjoyed by others.<sup>2</sup>

What has long lain under our eye, and has often been employ'd to our advantage, *that* we are always the most unwilling to part with; but can easily live without possessions, which we never have enjoy'd, and are not accustom'd to. 'Tis evident, therefore, that men wou'd easily acquiesce in this expedient, *that every one continue to enjoy what he is at present possess'd of*; and this is the reason, why they wou'd so naturally agree in preferring it. (T 3.2.3.4/503–04; italics original)

There are other possible rules, of course, for demarcating property. But the rule of present possession is the one that 'engages the attention most' (T 3.2.3.6/505). Thus, strategic rationality is what enables us to establish general rules for the sta-

ble possession of property, but it is the faculty of imagination which allows us to coordinate upon particular rules for determining exclusive use of these goods.<sup>3</sup>

### Social Exchange and Self-Control

Conventions for the stable possession of property are useful, but what if we find ourselves with a surplus of goods? In order to reap the full benefits of economic cooperation, according to Hume, we must settle upon conventions for social exchange.

Different parts of the earth produce different commodities; and not only so, but different men both are by nature fitted for different employments, and attain to greater perfection in any one, when they confine themselves to it alone. All this requires a mutual exchange and commerce (T 3.2.4.1/514; cf. T 3.2.2.3/485)

It would have occurred to our ancestors by 'plain utility and interest,' as he puts it, that they would mutually benefit from the transfer of surplus goods and services (T 3.2.4.2/515). Thus, we can say that agents enter into *social exchange conventions* if and only if:

- (A) Each agent believes that the other is strategically rational.
- (B) Each agent believes that social exchange would promote their individual interests.
- (C) Each agent believes that others believe that social exchange would promote their individual interests.
- (D) The beliefs in (A)–(C) produce a suitable regularity in their behaviour.

Social exchange conventions would have gradually evolved, then, as strategically rational agents came to acknowledge the reciprocal advantages of trade and commerce.

Hume recognizes that such conventions are difficult to explain, however, in the context of large-scale commercial societies. The problem is that the transfer of possessions in economies of any significant size will involve temporally deferred exchanges. Suppose that I promise to reimburse you next week for your goods or services today. You would not agree to the exchange unless you were assured that I will keep my word when the appropriate time comes. If we are complete strangers to each other, however, you would have no reason to believe this.

Your corn is ripe to-day; mine will be so to-morrow. 'Tis profitable for us both, that I shou'd labour with you to-day, and that you shou'd aid me to-morrow. I have no kindness for you, and know you have as little for me. I will not, therefore, take any pains upon your account; and should I labour with you upon my account, in expectation of a return, I know I shou'd be disappointed, and that I shou'd in vain depend upon your gratitude. Here then I leave you to labour alone: You treat me in the same manner.

The seasons change; and both of us lose our harvests for want of mutual confidence and security (T 3.2.5.8/520–21).

We can trust that our friends and family will keep up their end of the bargain, according to Hume, since they are naturally concerned about our welfare; but such confidence is much harder to come by when we are dealing with strangers who are largely indifferent to our plight.

Conventions for social exchange cannot be established unless strategically rational agents are assured that their partners will reciprocate. But how can they trust one another to do so? Hume proposes that strategically rational agents would recognize that it is in their *long-term* interests to cooperate in social exchanges. They would understand that anyone who breaks their word, as he puts it, 'must never expect to be trusted any more' (T 3.2.5.10/522).

Hence I learn to do a service to another, without bearing him any real kindness; because I foresee, that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me or with others. And accordingly, after I have serv'd him, and he is in possession of the advantage arising from my action, he is induc'd to perform his part, as foreseeing the consequences of his refusal. (T 3.2.5.9/521)

The crucial point is that agents stake their reputations on these exchanges, and thus failure to reciprocate would deprive them of the benefits of future commerce. Even though unilateral defection is a dominant strategy in a *one-shot* farmer's dilemma, in other words, this is not the case when it comes to an *iterated* sequence of games. Conventions for social exchange are established, therefore, because agents expect one another to act on their long-term interests.

Hume recognizes that appeal to our considered interests, however, is not sufficient to explain temporally delayed sequential exchanges. The problem is that human beings have a strong psychological propensity to discount the future.

[E]very thing, that is contiguous to us...commonly operates with more force than any object, that lies in a more distant and obscure light. Tho' we may be fully convinc'd, that the latter object excels the former, we are not able to regulate our actions by this judgment; but yield to the sollicitations of our passions, which always plead in favour of whatever is near and contiguous.

This is the reason why men so often act in contradiction to their known interest; and in particular why they prefer any trivial advantage, that is present, to the maintenance of order in society, which so much depends upon the observance of justice. (T 3.2.7.2–3/535)

Prudence counsels us, in other words, to reciprocate in social exchanges; but it seems that our tendency to prefer *small-rewards now over large-rewards later* would render us incapable of taking such advice.<sup>4</sup>

We seem to have fallen right back into the jaws of the farmer's dilemma. There would be no reason to cooperate in these circumstances because this propensity toward temporal discounting would become common knowledge, and as a result, strategically rational agents would expect each other to defect for the sake of short-term gains.

You have the same propension, that I have, in favour of what is contiguous above what is remote. You are, therefore, naturally carry'd to commit acts of injustice as well as I. Your example both pushes me forward in this way by imitation, and also affords me a new reason for any breach of equity, by showing me, that I should be the cully of my integrity, if I alone should impose on myself a severe restraint amidst the licentiousness of others. (T 3.2.7.3/535)

It appears that our analysis of social exchange conventions was incomplete. Agents must not only believe that their partners recognize the advantages of social exchange; they must also believe that their partners are *self-controlled* enough to act on their acknowledged interests. If human beings are incapable of delayed gratification, however, this condition could never be satisfied. Strategically rational agents would not enter into delayed exchanges with shortsighted and impulsive partners.

Hume recognizes that social exchange conventions could never have been established in large-scale societies if our tendency to discount the future was left unchecked.<sup>5</sup> In order to complete his natural history of justice, therefore, he must explain how our ancestors managed to overcome this obstacle. Hume rejects any naïve solution which suggests that agents can conquer their impulsivity through strenuous effort and a 'repeated resolution' to be strong-willed (T 3.2.7.5/536). The remedy must, he maintains, be significantly harsher. Indeed, the only way for shortsighted agents to acquire self-control is to submit themselves to the authority of civil government.

Men are not able to radically cure, either in themselves or others, that narrowness of soul, which makes them prefer the present to the remote. They cannot change their natures. All they can do is change their situation, and render the observance of justice the immediate interest of some particular persons, and its violation their more remote. These persons, then, are not only induc'd to observe those rules in their own conduct, but also to constrain others to a like regularity, and enforce the dictates of equity thro' the whole society. (T 3.2.7.6/537)

Government sanctions, in contemporary terms, serve as a 'commitment device'.<sup>6</sup> Like alcoholics who voluntarily check themselves into a clinic, we become aware of our propensity to discount the future, and thus commit ourselves to a coercive rule of law. Governments are not established, on this account, to shelter us from others; rather, they serve to protect us from ourselves.<sup>7</sup>

This institutional solution, however, cannot be the entire answer. For one thing, it is hard to understand why strategically rational agents would expect civil magistrates to act in a self-controlled manner. It seems the only way that they could avoid acting impulsively, after all, would be to submit to a higher coercive authority, and so on *ad infinitum*. Hume attempts to block this infinite regress by stipulating that the execution of justice lies in the *immediate* interests of the magistrates (T 3.2.7.6/537). One might wonder about the plausibility of this claim. There is a more fundamental worry, in any case, about the appeal to government as a commitment device. The problem is that the magistrates would presumably enforce social exchange conventions through punitive sanctions. But such threats would be ineffective deterrents for shortsighted and impulsive agents. If it is really the case that we discount the future, we would overlook any risk of future punishment.

### Chastity and Justice

We can find clues to a more promising solution in Hume's account of chastity norms. Hume's natural history of chastity begins with a number of *prima facie* plausible assumptions (T 3.2.12.3/570–71):

- (i.) Men and women have a common interest in raising children to maturity.
- (ii.) Human infants require lengthy and costly parental care.
- (iii.) Men prefer not to contribute to these costs unless the child is biologically related to them.
- (iv.) Men cannot, given the facts of human anatomy, be certain of paternity.

Men face an assurance problem, then, when it comes to raising children. They are unwilling to contribute to child care unless they are assured of paternity. But how can they be confident this is the case? How can they trust, in other words, that they have not been cuckolded?

It is obvious that vows of fidelity cannot solve this problem. The question is whether a particular agent is trustworthy. Thus, one cannot simply take them on their word. (In the language of game theory, this is 'cheap talk'.) Hume also considers, and quickly rejects, the suggestion that institutional sanctions could provide the requisite security. The problem is that 'legal proof', as he puts it, would be 'difficult to meet with in this subject' (T 3.2.12.4/571). This is not true in the court of public opinion, however, where the standards for indictment are notoriously low.

What restraint, therefore, shall we impose on women, in order to counter-balance so strong a temptation as they have to infidelity? There seems to be no restraint possible, but in the punishment of bad fame or reputation; a punishment, which has a mighty influence on the human mind, and at the same time is inflicted by the world upon

surmizes, and conjectures, and proofs, that would never be receiv'd in any court of judicature. (T 3.2.12.4/571)

Women have an interest in maintaining their reputation in society. But a name can be lost upon the slightest presumption of impropriety. They will carefully avoid any situations, then, that might provoke indignation from others. Men can rest assured that their mates, in turn, will avoid any temptation of infidelity: the risks would simply be too great.

This proposal does not, according to Hume, go far enough. Women have a propensity, like everyone else, to discount the future; as a result, they will naturally overlook the damage that might be done to their reputations.

All human creatures, especially the female sex, are apt to over-look remote motives in favour of any present temptation: The temptation is here the strongest imaginable: Its approaches are insensible and seducing: And a woman easily finds, or flatters herself she shall find, certain means of securing her reputation, and preventing all the pernicious consequences of her pleasures. (T 3.2.12.5/572-573)

It is not sufficient for men to believe that infidelity goes against the long-term interests of their mates. Even if women acknowledge their interests in fidelity, after all, they might be unable to resist the allure of the present moment. Thus, we find ourselves back where we started. Men will not contribute to the costs of child care unless they believe that their mates are self-controlled enough to act on their acknowledged interests. If our propensity toward temporal discounting is common knowledge, however, this necessary condition would never be satisfied.

Hume maintains that this assurance problem could not be solved unless men believe that their partners are disposed to feel repugnance at the very thought of infidelity.

'Tis necessary, therefore, that, beside the infamy attending such licenses, there should be some preceding backwardness or dread, which may prevent their first approaches, and may give the female sex a repugnance to all expressions, and postures, and liberties, that have an immediate relation to that enjoyment. (T 3.2.12.5/572)

It is not enough for women to believe, in other words, that they would feel ashamed if their infidelity is revealed. These concerns about the future will be discounted and thus cannot override the enticement of immediate pleasures. These negative affects must occur at the moment of deliberation.<sup>8</sup> Agents act impulsively because they overestimate the value of present goods. The only way this propensity could be corrected, then, is if these rewards were rendered less attractive. And this is precisely what is accomplished by the feelings of backwardness and dread: they serve as contrary hedonic impulses.

Hume recognizes that he must account for these emotional aversions. It is *prima facie* implausible to maintain that women would feel repugnance, after

all, at the 'approaches of a pleasure, to which nature has inspir'd so strong a propensity' (T 3.2.12.6/572). He maintains that this aversion can be explained, however, in terms of a mixture of sympathy and socialization. Those who have an interest at stake will *naturally* disapprove of female infidelity; but even those who lack such concerns will *morally* condemn any signs of immodesty because they are 'apt to be affected with sympathy for the general interests of society' (T 3.2.12.7/572). These chastity norms are subsequently reinforced by educators who teach women, from an early age, to look upon any impudence with a wary eye. Public censure and private education work together to insure, then, that women become disinclined toward an otherwise natural impulse.

Hume's natural history of chastity points to an alternative account of how our ancestors managed to establish conventions for delayed social exchanges. The challenge is to explain why strategically rational agents would trust their shortsighted partners to reciprocate once they have received the benefits of the exchange. The solution is that agents would not participate in these sequential exchanges unless they had reason to believe that their partners are disposed to feel repugnance at the very idea of cheating. Mutual trust is made possible, in other words, by our 'abhorrence of injustice' (T 3.2.2.25/500). It is the fact that you are affectively repulsed by the thought of cheating which assures me that you are not only prudent enough to recognize the advantages of reciprocal exchange, but you are also self-controlled enough to act on your acknowledged interests.

Hume recognizes that he is obligated to account for this antipathy toward injustice. How could impulsive agents become disinclined toward cheating, especially when unilateral defection offers such a large short-term payoff? Hume's answer proceeds along the same lines as his discussion of chastity norms. Those with interests in a particular exchange will naturally disapprove of defectors, and even those without anything at stake will 'partake of their uneasiness by *sympathy*' (T 3.2.2.24/499).<sup>9</sup> This aversion to injustice is subsequently reinforced, as with chastity norms, by private education and public rhetoric (T 3.2.2.25-26/500-01).

This interpretation throws new light on Hume's remarks about the sensible knave. Sensible knaves acknowledge that society could not exist without rules of justice, but they think it best to opportunistically violate them.

That *honesty is the best policy*, may be a good general rule; but is liable to many exceptions: And he, it may, perhaps, be thought, conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions. (EPM 9.22/282-3)

Hume argues that we need not worry excessively, however, about these deviant characters. Sympathy and socialization will insure that most of us acquire an 'antipathy to treachery and roguery' (EPM 9.23/283). What about those who

fail to develop these aversions? Hume concedes that such knaves exist, but he denies that they would be sensible enough to avoid detection.

[W]hile they purpose to cheat with moderation and secrecy, a tempting incident occurs, nature is frail, and they give into the snare; whence they can never extricate themselves, without a total loss of reputation, and the forfeiture of all future trust and confidence with mankind. (EPM 9.24/283)

Knaves are those who lack feelings of repugnance towards cheating. But these emotions are required, as we have seen, for human beings to achieve self-control. Those who lack an antipathy to injustice, then, will inevitably succumb to the temptation to cheat, even when it is unsafe to do so.

### Collective Action

It is unfortunate that Hume did not pursue this sentimentalist approach to justice, since it would have enabled him to solve another important puzzle about human cooperation. Even if strategically rational agents were self-controlled, Hume recognizes, there would be many circumstances in which they could not join forces. His meadow example illustrates one such situation.

Two neighbours may agree to drain a meadow, which they possess in common; because 'tis easy for them to know each other's mind; and each must perceive, that the immediate consequence of his failing in his part, is, the abandoning of the whole project. But 'tis very difficult, and indeed impossible, that a thousand persons shou'd agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expence, and would lay the whole burden on others. (T 3.2.7.8/538-9)

The draining of the meadow is a collective action problem.<sup>10</sup> The difficulty is that each person would reap the benefits of the project, whether or not they help out with its costs. Agents would recognize that they could free-ride upon the efforts of others, therefore, and nobody would make any contributions.

Let us define a *collective action convention* in the following terms. Individuals enter into collective action conventions if and only if:

- (A) Each agent believes that the other is strategically rational.
- (B) Each agent believes that contributing to public goods would promote their individual interests.
- (C) Each agent believes that others believe that contributing to public goods would promote their individual interests.
- (D) The beliefs in (A)–(C) produce a suitable regularity in their behaviour.

The problem is that it is difficult to see how these conditions could ever be satisfied. Individuals would not believe that it was in their interests to contribute to

public goods, nor would they expect others to arrive at this conclusion. In short, it appears that collective action would be impossible for strategically rational agents. The worry in this case is not that agents are too impulsive to act on their interests; rather, contributing to public goods was never in their interests in the first place.

This problem does not arise when group size is small, according to Hume, because agents can monitor each other's behaviour. It is much harder to keep track of free-riders, however, in the context of large-scale societies. How could our ancestors have established collective action conventions in these circumstances? Hume once again invokes governmental sanctions. The meadow never gets drained because agents prefer to let others do all the work. But if institutions were established that detect and punish free-riders, agents would be compelled to contribute to public works.

Thus bridges are built; harbours open'd; ramparts rais'd; canals form'd; fleets equip'd; and armies disciplin'd; every where, by the care of government, which, tho' compos'd of men subject to all human infirmities, becomes, by one of the finest and most subtle inventions imaginable, a composition, that is, in some measure, exempted from all these infirmities. (T 3.2.7.8/538-9)

The threat of punishment would create an incentive for agents to do their part. Governments would *make* it their interest, as it were, to contribute their fair share.

Hume's institutional solution once again fails, however, to provide a sufficient remedy. The central difficulty is that the invention of government is itself a collective action problem.<sup>11</sup> Strategically rational agents would prefer that others contribute the resources required to create agencies for monitoring and enforcing participation in public works; they would prefer to reap the benefits of regulation, in other words, without incurring any of its start-up costs.<sup>12</sup> This would not be the case for individuals, however, who have acquired feelings of repugnance to injustice. Even though such agents lack *economic* incentives to participate in collective action, these affections would insure that they have *psychological* interests in contributing their fair share. As with social exchange conventions, then, emotional aversions allow us to overcome dilemmas that prudential reasoning alone could not.

### Conclusion: The Current Status of Hume's Enlightenment Project

The natural history of man was a central concern of the Scottish Enlightenment.<sup>13</sup> But Hume anticipates that these accounts might be brushed aside as nothing but 'chimerical speculation' (T 3.2.12.6/572). Such proposals refer to events, after all, beyond our written records. Can they be substantiated? Or are they entirely *conjectural*? Hume was not in a position, of course, to evaluate his account. But it is no longer necessary to remain agnostic about its status.

Researchers in experimental game theory and neuroeconomics have developed a novel paradigm that allows us to put it to the test.

This experimental approach is admittedly in its early stages, but the results are coming in, and they are consistent with Hume's solution to the puzzle of cooperation. Researchers have discovered that negative reward circuits in the brain become active, for example, when participants make uncooperative moves in games involving social exchange.<sup>14</sup> It appears that this circuitry enables participants to cooperate in these games, moreover, by inhibiting their impulsive desires for immediate gratification.<sup>15</sup> Further studies have revealed that 'negative emotions' also serve as the affective mechanisms which ensure that participants contribute their fair share in public goods games.<sup>16</sup>

There is a growing consensus in cognitive science that emotions play an important role in economic behaviour. The causal etiology of these affections, however, remains an open question. It is unclear to what extent, for example, they are innate or learned. Hume's hypothesis is that our aversions to injustice result from our automatic tendency to resonate with the affections of those around us. There is some preliminary support for this proposal. Recent studies have shown that areas of the brain associated with vicariously sharing the emotions of others, for example, become active when participants make uncooperative moves in iterated prisoner's dilemmas.<sup>17</sup> This hypothesis, of course, requires further testing. But the crucial point is that it could provide an important research program for those interested in the puzzle of why human beings cooperate with one another to the unique extent that we do.

Previous commentators have correctly emphasized the role of game-theoretic notions in Hume's theory of justice.<sup>18</sup> Prudential considerations clearly play a pivotal role in this account. But Hume recognizes that they cannot fully explain how creatures such as ourselves – with our propensities to discount the future and free-ride on the efforts of others – could have established justice conventions in large-scale societies of strangers. This leads him to present an innovative and plausible account of the psychological prerequisites of conventions: human beings manage to cooperate with one another because we are *strategically rational creatures with a heart*.