

# Rationality and Society

<http://rss.sagepub.com/>

---

## **Feeling cooperation, being moral**

Andrés de Francisco

*Rationality and Society* 2014 26: 355

DOI: 10.1177/1043463113512998

The online version of this article can be found at:

<http://rss.sagepub.com/content/26/3/355>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Rationality and Society* can be found at:**

**Email Alerts:** <http://rss.sagepub.com/cgi/alerts>

**Subscriptions:** <http://rss.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://rss.sagepub.com/content/26/3/355.refs.html>

>> [Version of Record](#) - Jul 22, 2014

[What is This?](#)

# Feeling cooperation, being moral

Rationality and Society  
2014, Vol. 26(3) 355–372

© The Author(s) 2014

Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/1043463113512998

[rss.sagepub.com](http://rss.sagepub.com)



**Andrés de Francisco**

Complutense University, Madrid, Spain

## Abstract

Is morality a pure rational abstraction or is it rooted in human nature? In this paper I argue for the second option, and I do so in several steps. I first vindicate the cognitive role of emotions. For this, I rely on two contemporary theories of emotions, Martha Nussbaum's and Antonio Damasio's, without forgetting the evolutionary approach to emotions. Then I defend a complex model of human behaviour that goes beyond rational choice models and includes, following Gintis, other-regarding orientations and altruistic motivations. Key ingredients of such a complex model are the social emotions involved in human cooperation and reciprocal altruism. Underlying social cooperation is an emotional economy which turns out to be highly informative about the moral nature of social life. I analyse this emotional economy of cooperation and, following Aristotle, I show its moral foundations. Finally, I claim that the set of moral norms and the corresponding emotional responses involved in cooperative behaviour could be reduced to a unique moral principle. Here I draw on the general moral principle – *neminem laede; imo omnes, quantum potes, juva* – and follow Schopenhauer in his philosophical derivation of the principle of compassion from the emotion.

## Keywords

Compassion, cooperation, morality, social emotions

The Roman Stoic Seneca said that 'the passions are as bad servants as they are leaders'.<sup>1</sup> Mainstream western philosophy has been more nuanced but,

---

## Corresponding author:

Andrés de Francisco, Departamento de Sociología I, Facultad de CC. Políticas y Sociología, Universidad Complutense de Madrid, Campus de Somosaguas, Madrid 28223, Spain.

Email: [afrancis@ucm.es](mailto:afrancis@ucm.es)

with some remarkable exceptions, has considered emotions mainly as dangerous impulses that should be controlled or even enslaved by reason if we are to have a good life.<sup>2</sup> The rational choice theoretical framework – once dominant in social sciences, and still highly relevant – went so far as to completely overlook emotions. In its formal models, beliefs and desires are taken as given, but emotions are not even mentioned; they simply do not exist.

But emotions do exist. Indeed, in recent years a considerable body of outstanding research, conducted by evolutionary psychologists and neuroscientists, by paleoanthropologists and experimental economists, by sociologists and cognitive psychologists, as well as by philosophers of the mind and moral philosophers, has challenged both the reductionist framework of rational choice theory and the secondary role historically assigned to emotions, and proposes a very different view of human behaviour that includes at least the following theses: 1. Emotions are not just the irrational affective component of human motivation but have decisive cognitive and evaluative dimensions necessary for a good and meaningful life. 2. Some of our social emotions have an ethical underpinning that has to do with reciprocity and fairness, two basic features of human cooperation. In this paper I defend the idea that cooperative behaviour cannot be fully understood if we abstract from (a) the cognitive-evaluative role played by emotions; and (b) the relations between social emotions and moral judgement. Additionally, I show that the emotional economy underlying cooperative behaviour (based on reciprocity and trust) is crucially informative about the moral nature of human agency. I close the paper defending the following hypothesis: that the rich variety of moralistic emotions of cooperation can be reduced to a general and unique moral principle.

## **On the cognitive nature of emotions**

In recent decades, there has been increasing academic and scientific interest in the nature of emotions and their role in human behaviour. Most scholars agree that emotions include (or presuppose) at least beliefs and evaluations, although they are not reducible to them, since they also incorporate other dimensions such as desires, sensations, feelings, physiological processes and behavioural traits. Whether the cognitive elements of emotions are best captured by a propositional, judgemental, computational or even perceptual framework of analysis is a question of ongoing philosophical debate.<sup>3</sup> It goes without saying that cognition is not rationality. In fact, the beliefs underlying emotions may be irrational (ill-formed) beliefs and there may be inadequacy or lack of proportion between the stimuli and the emotional reaction, whose direct consequences and side-effects may jeopardize

practical rationality. Nevertheless the frontier between emotions, cognition and rationality is anything but rigid – as it used to be in mainstream western philosophy – logocentric and ‘Apollonian’. With brief references to some other contributions, here I will start by presenting two different approaches, coming from philosophy and neuroscience respectively, which highlight the *intelligence* of emotions. These are Martha Nussbaum’s neo-Stoic theory of emotions and Antonio Damasio’s insights into the cognitive value of emotions for rational decision-making under uncertainty.

Nussbaum (2001: 19) synthesizes her view of emotions as follows:

Emotions... involve judgements about important things, judgments in which, appraising an external object as salient for our own well-being, we acknowledge our own neediness and incompleteness before parts of the world that we do not fully control.

My fear, my hope, my love, my grief in this sense embody a ‘way of seeing’ (Nussbaum, 2001: 27) which is intentionally directed toward an external object. This intentional way of seeing invests the object with characteristics that place it in a special relation to my own well-being, my own flourishing and happiness. Emotions thus incorporate value judgements through which we understand the world, an understanding that reflects our own vulnerability with respect to things and persons out there that, being important to us, we do not fully control. I see a dead body and I could give a cold factual description of the matter of fact: position, location, dimensions, temperature, colour of skin, etc. This may be the way in which the forensic expert depicts the body, but is not at all the way a father sees his son’s dead body. For this loving and grieving father, that body is invested with a set of features related to memories that make up a shared past experience with his son. Seeing that beloved body lying there is seeing his own loss and his own vulnerability: nothing can be done about the tragedy. Emotions introduce the narrative of our lives into the cognitive interpretation of the world. Through emotions we *appraise*<sup>4</sup> the objects in the temporal sequence of our own experience, bringing past memories into the present evaluation of the object and projecting it into the future with our expectations, fears and hopes.

But emotions not only determine the way we see the world. Emotions also provide our own beliefs with a sort of truth-value test. For example, I will only be sure that I *truly* love that person when I have the relevant emotions: fear of losing her, for instance, and the corresponding feeling of my own incompleteness. Otherwise I may believe that I love her, that she is the correct person, but without those emotions I will just have *false* beliefs about that person and my emotional relation to her. Because

emotions are cognitive devices intentionally directed toward an object with subjective value, emotions are concerned with the person's flourishing and well-being, and thus allow for a *eudemonistic* interpretation: emotions seem to be a fundamental part of the ethics of happiness. In fact, important experiences of life which define us as flourishing human beings – friendship and love, cooperation and solidarity, self-esteem and self-realization – could not be held without the correct emotion that places the relevant set of values on the object, but altogether exposes us to frustration and loss. In short, emotions make us *see* the world in a certain way, in a special welfare relation to ourselves; guide our value judgements about external objects, determining our understanding of the world and our own experience of it; and finally regulate the truth-value of our beliefs. Far from just being a distorting factor of human mental processes and an independent source of irrational motivation (which of course they can also be), emotions play a deep and decisive cognitive and ethical role in human life. So far, Nussbaum's theory.

In spite of his being an eminent neuroscientist, Damasio has developed a fascinating *cognitive* theory of emotions and feelings based on neural and mental maps of body states and responses. But here I want to refer to his no less fascinating explanation of the cognitive role of emotions, especially in decision-making under uncertainty. The theory is based on research undertaken with patients who suffer brain damage of the ventromedial sector of the frontal lobe. Damasio and other neuroscientists found that these patients have no cognitive deficiencies and can solve logical problems but 'are emotionally flat at the level of their social emotions... such as embarrassment, sympathy and guilt' (Damasio, 2003: 144). This flatness seems to be related to their inability to make correct decisions under uncertainty and condemns them to myopia of the future, which in turn leads to their loss of social independence. Why is this so? Through which mechanisms does the causal process take place? The basic mechanism is the *revival of emotional signals*. This is possible because we humans have an emotion-related memory that we can activate in the appropriate circumstances. In Damasio's (2003: 146–147) own words, the whole argument goes like this:

Under the influence of social emotions (from sympathy and shame, to pride and indignation) and of those emotions that are induced by punishment and reward (variants of sorrow and joy), we gradually *categorize the situation we experience* – the structure of the scenarios, their components, their significance in terms of our personal narrative... When a situation that fits the profile of a certain category is *revisited* in our experience, we rapidly and automatically deploy the appropriate emotion.

This revived emotional signal marks the options and outcomes of the situation with a positive or negative value, anticipating the consequences of action, narrowing the decision-making space and increasing the probability that the action will conform to past experience (Damasio, 2003: 148). Thus, emotions stocked in our memory become kinds of harbingers of the future. This set of ideas is referred to as ‘the somatic-marker hypothesis’. Perhaps an example will help. Imagine we face a situation that can make us feel ashamed or embarrassed. An emotionally healthy individual, with his frontal lobe intact, will automatically search for the correct emotions in the system’s history of his life experience – that is, the past emotions that fit into the present categorization of the situation. These emotions mark the situation with negative signals and induce action that avoids the bad consequences: perhaps making the subject stay quiet and silent, or making him go away or get dressed in a different way, etc. Damasio’s argument narrowly resembles De Sousa’s (1987: 181–184) notions of *paradigm scenarios* and *emotional repertoires*. In both cases, there is a biographical history of emotional socialization and learning, guided by punishment and reward, that helps individuals develop, from a very early stage in their lives, a set of more or less automatic responses to contexts categorized in a certain way. These ideas can be easily translated into phylogenetic terms, taking into account not only the particular individual biographical narrative and her cultural environment, but the whole adaptive history of the species. In fact, that is precisely what evolutionary psychology has accomplished. From this point of view, not only primary emotions, such as fear and anger, but also social emotions, such as shame, guilt, gratitude or indignation – which are far more complex – are evolved sets of emotional programmes designed to respond functionally to ‘evolutionary recurrent situations’. Those programmes consist of complex sets of evolved *algorithms* for detecting the situation, reading the relevant cues and assigning priorities, while the correct application of those algorithms presupposes the activation of sophisticated cognitive processes involving perception, goal definition, memory, information gathering, attention programmes, instructions to the body and communication with other individuals through emotional expressions (Cosmides and Tooby, 2000). We shall come back to this research programme in the next section.

## **Cooperation, rationality and the social emotions**

It is certain that rational choice, on the assumption of self-regarding motivations and Bayesian rationality, has demonstrated the analytical possibility of endogenously evolved cooperation with repeated game models ‘in which punishment of defectors by cooperators secures cooperation among

self-regarding individuals' (Gintis, 2009: 181). There is even the so-called *folk theorem*, which states that, under certain conditions, there is a range of Nash equilibria for any iterated game repeated an indefinite number of times. Gintis (2009: 195), a firm supporter of game theory and clearly convinced of the analytical relevance of the folk theorem, nevertheless points at its central weakness, to wit: it pays no attention to 'how the Nash equilibria whose existence it demonstrates can actually be instantiated as a social process'. Moreover, the conditions under which the theorem is correct turn out to be extremely unrealistic. In fact, Gintis (2009: 196) continues, 'the folk theorem with self-regarding agents fails when agents are present-oriented, signals are imperfect, or players are likely to err' – that is, under conditions of real social life. To fully understand social cooperation, Gintis concludes, we need to complement rational choice theory with a psychological model of social preferences and a social epistemology of social norms, where the assumption of self-regarding motivations gives way to a more complex model of human behaviour which includes other-regarding orientations and altruistic motivations. Undoubtedly, social emotions are a key ingredient of any such socio-psychological model of social cooperation.

The deep connection between human emotional life and the adaptive functionality of social cooperation has been demonstrated by numerous observations in the fields of primatology and evolutionary psychology. Next I will briefly give just a few relevant examples of these contributions. For the eminent primatologist Frans de Waal, very simply, human beings have mind but also body. The body suffers because it is limited and vulnerable. As with other primates, the body turns us into inherently needy beings. Because of that, we live in mutual dependence. This dependence takes place inside social relations regulated by a golden rule – reciprocity – which implies, among other things, a sense of justice and the corresponding moral emotions (De Waal, 2006). The theoretical psychologist and primatologist Humphrey (2002) has brilliantly shown that the complex culture of cooperation has forced humans to develop not only a rich emotional life but also a complex cognitive machinery for understanding others' emotions, which implies an 'inner eye' – that is, self-consciousness: a fact that turns us into 'natural psychologists'. From the point of view of evolutionary psychology, Cosmides and Tooby (2000: 103–104) openly link Ekman's well-established theory of cross-cultural, universal, facial expressions of human emotions with the social ecology of human cooperation. Indeed, they argue, the fact that many of our emotional expressions are automatic and involuntary signals that broadcast our inner feelings provided our ancestors with an adaptive advantage, *because* our ancestors were cooperators for whom reciprocal altruism proved to be an essential ingredient of selective fitness.

From an evolutionary psychology perspective, in other words, the critical adaptive importance of social cooperation and reciprocal altruism in species evolution seems to explain the presence of this automatic emotion signaling. Not only that, but our whole cognitive architecture has co-evolved with our emotional repertoires. As Steven Pinker (1997: 405) puts it, ‘the human brain was driven by a cognitive arms race, fuelled by the emotions needed to regulate reciprocal altruism’.

Reciprocal altruism is indeed very demanding of cognitive mechanisms. In particular, it implies memory, recognition and cognitive discrimination. Especially in order to discriminate between cheaters and non-cheaters, false cooperators and true cooperators, it is crucial for the evolution and stabilization of cooperation. But these cognitive strategies seem to be firmly rooted in our genetic infrastructure, as two theories have demonstrated. First, Tooby and Cosmides’ *social contract theory* has proved that we spontaneously apply a logical rule – *modus tolendo tolens* – to detect the cheater in a cooperative context, without being conscious of it and without even understanding the logical structure of the rule. But we do use the rule, and it efficaciously works to detect the cheater. Apart from that, it has been shown by Trivers and others that we have fine-tuned cognitive mechanisms to discriminate between *sham emotions* and *real emotions*, something which has allowed the evolution of *trust* among cooperative agents. To this festival of evidence-gathering in favour of the cooperative nature of human sociability has recently been added behavioural game theory. In particular, experimental games, such as the Dictator Game, the Ultimatum Game or the Public Good Game, have demonstrated that humans are *strong reciprocators*, which implies a dual disposition for altruistic cooperation (in the face of social dilemmas) and for altruistic punishment when reciprocity is not obtained. It can be said that without this second disposition – our willingness to *punish* the non-reciprocator – cooperation could never have been a stable strategy. Note that altruistic punishment means ‘hurting others at a cost to oneself’, so it does not seem to be a rational self-regarding strategy. But we actually are prone to punish cheaters or norm violators even at a cost to ourselves. That this is a quite natural predisposition also anchored in our genome is clearly pointed out by the neurological discovery of the pleasure effect – located in the dorsal striatum – of righteous punishment (DeQuervain et al., 2004).

In conclusion, we can say that the centrality of cooperation and reciprocal altruism in human-evolved nature, as well as the functional centrality of emotions in cooperation, is well supported by empirical evidence and theoretical construction. The structure of human cooperation leads us beyond pure rationality into the field of emotions and feelings.



## Feeling cooperation

But what do we feel while cooperating? Which emotions are involved in strong reciprocity? Is there an emotional economy underlying cooperation?

Before trying to answer, two caveats are in order. First, there is Jon Elster's (1999: 48–50) warning on the intrinsic methodological limitations of the analysis of emotions, due to objective limitations in evidence-gathering and direct observation. Emotions are not things out there to be seen and measured. That is the reason why introspective plausibility and acknowledgement are unavoidable where emotions are concerned. Perhaps this is not a limitation at all but a methodological advantage: as Descartes wrote long ago, 'everyone has experience of the passions within himself, and there is no necessity to borrow one's observations from elsewhere in order to discover their nature'.<sup>5</sup> Second caveat: cooperation is either a joint effort aimed at achieving a common objective or a tit-for-tat iterated game. They rarely, if ever, take place in a social and historical vacuum; on the contrary, cooperation is a social relation with a history, which means that it is determined by the agents' memory of past sequences of episodes of successful or failed reciprocity, and it is directed at future movements: the cooperative agent looks back at the past for information but also enters the future with a set of expectations. Therefore, trust – a central 'externality' of social interaction that 'lubricates' the social system (Arrow, 1974: 23) – becomes a crucial feature behind the decision to cooperate. What I have to say here about trust and trustworthiness follows Russell Hardin's 'encapsulated interest account' of trust, but is compatible with the other two great theories of trust (Hardin, 2006: ch. 2).<sup>6</sup> Here I will abstract from questions of power or hierarchy, and will only consider the case of *horizontal cooperation among free equal agents*.

For a better answer to the above questions, I will separate two different moments in the cooperation process: the *ex-ante* situation in which agents decide whether or not to cooperate, and the *ex-post* situation in which agents decide to continue cooperating or to punish.

### *Ex-ante cooperation*

1. *A cooperates*. To cooperate involves the risk of being cheated or free-ridden. Before risking an action that will not be paid back, agent *A* wants to make sure that agent *B* will do her part. If the answer is positive, *A* cooperates trustfully. Trust is an important non-market commodity – an externality, as I said – but also a complex state of the mind. It means that I *expect with high subjective probability* that you will do your part and reciprocate, presently or in

the future. Virtually all conceptions of trust include an element of expectation (Hardin, 2006: 29). On the other hand, without that expectation of reciprocity, it is difficult to imagine the evolution of cooperation. Hence it is no surprise that, translated into the language of emotions, trust (and cooperation) involves *hope*, a future-oriented emotional state of the mind.

Nevertheless, hope does not fully explain the motivation of the first movement of giving. If it is true that humans are strong reciprocators, this first movement must be guided by a different disposition. This disposition could be assimilated to what Aristotle called *charis*: spontaneously doing a favour for the person in need.<sup>7</sup> Such spontaneous generosity of character is probably anchored in natural sympathy, which is an open-ranged emotion oriented to the emotional states of other people. Through sympathy, in fact, the utility functions of different individuals get connected and become linearly dependent on each other. As long as your utility is also mine, it is natural that I care for you. But such caring is impossible without sympathetic feelings.<sup>8</sup>

Nevertheless, in spite of the genuine generosity characteristic of the first movement in strong reciprocity, this does not eliminate the underlying expectation of getting the favour back when the occasion presents itself. Generosity demands gratitude, and disappointment typically follows the lack of reciprocity: if you do not get the favour in return, it is unlikely or less probable that you will show *charis* again with the non-reciprocating person on a future occasion.

In general terms, though, cooperation is not a 'charitable' enterprise. It is rather a strategic game where one gives and takes, a game of interchanging effort, as de Waal puts it: a 'market of services'. And here the expectation of reciprocity is crucial. The 'encapsulated interest account' of trust and trustworthiness grounds the positive expectation (confidence) of the truster in the trusted's own interest to act cooperatively. But even so, as Hardin acknowledges, trustworthiness is never completely guaranteed, as the trusted person might have competing interests that trump the truster's. So, there is always some risk in acting on trust (Hardin, 2006: 26). That is why hope has a complementary partner in the emotional economy of trust and cooperation: fear. When you hope for reciprocity, you also fear that it will not occur. The stronger the hope, the weaker the fear. Trust increases linearly with the first emotion and decreases with the second. That is why in real friendship (at least ideally)<sup>9</sup> there is no fear but plenty of hope and trust. But although friendship is the most cooperative of social relations, social cooperation is not so friendly, and in most cases incorporates a certain level of fear or, if you want, anxiety; this is because social cooperation takes place

in the realm of relative uncertainty. And that is why, in line with Damasio's theory, these emotions are so important: they help manage the *market of services* in a context of relatively uncertain reciprocity. I dare say that, without a functional amount of fear or anxiety, cheaters and free-riders will dominate and cooperation will fail to stabilize. Fear keeps us awake and alert; fear supports our cognitive search for would-be defectors. It makes us suspicious. Of course, the emotional partner is also crucial: without an even greater amount of hope, cooperation would never become a stable strategy either. So the emotional basis of ex-ante cooperation seems to be a certain combination of *hope* and *fear*,<sup>10</sup> whose relative weights will depend on the level of interpersonal trust so far accumulated within the particular social relation or within social relations generally.

But why should we hope for reciprocity in cooperation? Out of mere interest? We are interested in reciprocity, of course, because cooperation yields clear benefits, be they public goods or gratitude and altruistic returns. These goods and returns of *social capital*, in line with Nussbaum's eudaimonistic approach to emotions, are part of our well-being and should be included in any definition of a good human life. But is that the whole story? Or could there also be ethical components necessarily implied? Generosity calls for generosity: that is a fact of human sociability. Invariably, in fact, we cannot avoid disappointment when we are not paid in kind, just as we cannot help feeling satisfied in the face of reciprocated altruism or gratitude.

The question is whether the emotional programmes regulating those responses –disappointment and satisfaction – are reducible to mere material self-interest or include some other dimension. In my view, they are also linked with morality. In this sense, entering or giving up social cooperation would incorporate a moral 'way of seeing' based on normative judgements, in particular the following: (a) giving and receiving proportionately and sharing efforts in a joint venture is not only a mutually beneficial equilibrium but also a *fair* equilibrium; and (b) the contrary is a breakdown of 'distributive' justice, not only of social efficiency. Moreover, these normative judgements would be related to our conception of a good life – 'being fairly treated is part of my well-being; distributive justice is essential for a well-ordered society' – and so would allow, in line with Nussbaum's theory, for a eudaimonistic interpretation.

Perhaps the best evidence of the link between moral judgement and reciprocity rests on our willingness to *punish* cheaters; in other words, perhaps the moral sense of justice in cooperation is manifested especially in the reactive response of punishment. Thus, we now want to explore the emotions involved in punishing non-cooperators.

## Ex-post cooperation

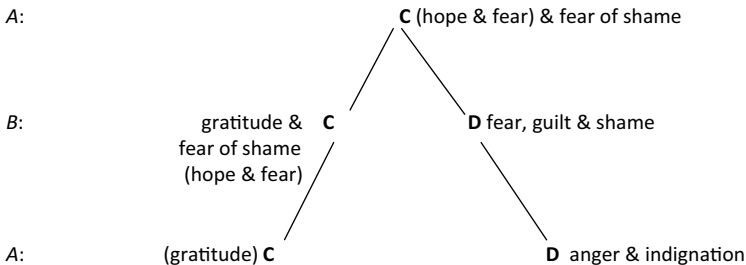
2. *B defects*. If *B* defects, the whole machinery of punishment starts working. The emotions involved are probably, on the part of *A*, a certain combination of *anger and indignation*. Following Aristotle, anger is ‘an impulse, accompanied by pain, to a conspicuous revenge for a conspicuous slight directed without justification towards what concerns oneself or towards what concerns one’s friends’.<sup>11</sup> Two ideas stand out here: the idea of *slight* and the idea of *desert (justification)*. Much before the discovery of the dorsal striatum, Aristotle was rather subtle when he added that anger ‘must always be attended by a certain pleasure – that which arises from the expectation of revenge’.<sup>12</sup> The will to punish (revenge) is implicit in the experience of anger, as it is the *pleasure* felt in imagining and anticipating it. Nevertheless, the will to punish and the *ex-ante* pleasure effect need not finish in actual punishment, as there might be some impediments – lack of resources, instrumental rationality – that frustrate the execution of the will to punish. The other emotion that *A* might feel is indignation. Aristotle defines indignation, in contrast to compassion, as ‘pain at unmerited good fortune’.<sup>13</sup> As we can see, both emotions – anger and indignation – refer to the value component of the *justice of merit*, ‘... for whatever is undeserved is unjust, and that is why we ascribe indignation even to the gods’.<sup>14</sup> Being cheated is being badly treated without justification, undeservedly. So we feel slighted and want to punish the offender. In addition, while despising our generous effort, the cheater takes unfair advantage of it, obtaining a benefit that he does not deserve, which motivates our indignation. Aristotle does not explore the behavioural side of indignation, but it is not difficult to guess. Indignation motivates action – mainly, protest – to redress an unfair distribution of advantages and reestablish a just correspondence to merit.

On the part of *B*, the cheater, what are the emotions involved? At least these: *guilt, fear* of being discovered (and punished) and, if so, *shame*. Aristotle relates shame to honour and says that shame is a ‘mental picture of dishonour (*phantasia adoxias: Rhetoric*, II, 7, 1384a)’ in which we represent the bad opinions others hold about us. That shame is accompanied by a feeling of pain or disturbance is due to two things. First, those persons matter to us; second, their bad opinions are justified in our own vices. Of course we feel ashamed even if (we know that) their bad opinions are wrong: the mere attribution of vices by persons who matter to us, even if undeserved, is a painful ‘phantasia’.<sup>15</sup> Vices are bearable – if they

are bearable at all – in intimacy or privacy. We do not want to show them in public. Why? Probably because there is no public space without norms of decency and moral rules, which define correctness, propriety and virtue. If we could display publicly our private vices without impediment, society would most likely be impossible. In a well-known passage of *The Laws*, Plato stresses the importance of educating the young citizen in the virtue of modesty or decency (*aido*) – that is, of inculcating in him the fear of being shameless (*anaischyntos*).<sup>16</sup> Indeed, shame is such a core emotion underlying any scheme of social norms that some scholars even think it arises whenever there is a threat to the *social bond* (Scheff, 2000: 95 ff.). In any case, it seems clear that shame or *fear of shame itself* should be included in any explanation of cooperation, as cooperation is an enterprise based on norms of fairness and reciprocity.<sup>17</sup> Many vices are involved in opportunistic and non-reciprocating behaviour, from incontinence to greed and lack of gratitude, but there is one that is of special importance here: injustice. The material of shame in this case is the *moral* reputation of the agent, not just his reputation as an agent capable of honouring conventional or social norms in general: I took *unfair* advantage, which is a violation of a specific norm of fairness; I did not return the *favour*, which is a violation of a specific norm of reciprocity; they shall think I am a *mean* person (I do not possess a noble character). It is this moral judgement which really hurts the ashamed person who chooses to default on cooperation.

3. *B cooperates*: *B* feels gratitude – or fear of shame – and reciprocates. But there remains an emotional combination of hope and fear, with hope (and serenity) being tendentially dominant.
4. *A cooperates after B has cooperated*. The same as 3.

The above distribution of emotions can be summarized as follows, where *C* represents cooperation and *D* defection



## The moral basis of cooperation: A reductionist hypothesis

As we have seen, the social emotions involved in cooperation ultimately refer to the moral sense of agents. Cheaters and non-reciprocators violate well-established (probably evolved) norms of morality and trigger the corresponding moralistic emotional responses. Do these responses form an unorganized set of responses, or is there a general moral principle to which all those moralistic emotions – anger, indignation, guilt, shame, gratitude – ultimately refer? In my opinion, the German philosopher Arthur Schopenhauer offers an interesting argument that supports a reductionist hypothesis.

In *On the Basis of Morality*, Schopenhauer (1995: 69) argues that there is a general moral principle, a principle ‘whose establishment is the constant endeavor of all teachers of morals’. This is: *neminem laede; imo omnes, quantum potes, juva* [‘Injure no one; on the contrary, help everyone as much as you can’]. Incorporating the golden rule of reciprocity [*Quod tibi fieri non vis, alteri ne feceris*],<sup>18</sup> this is the principle – so Schopenhauer thinks – that ultimately underlies all actions with *moral value*, that is, actions beyond self-interest or egoism.<sup>19</sup> The principle is expressed as a commandment, but it contains two differentiated moral judgements. The first states that injuring someone is *prima facie* a bad thing, so that it would need a special justification to be morally acceptable. The second moral judgement involved in the principle is that helping others is *prima facie* a good thing, so it would need a special justification to be morally unacceptable. That is why the universal moral principle contains two maxims: one impels me to do no harm, the other to do good.<sup>20</sup> These two different kinds of actions covered by the moral principle allow Schopenhauer to talk about two different *virtues* regulating them: the virtue of justice (*Gerechtigkeit*) and the virtue of philanthropy (*Menschenliebe*), which Schopenhauer (1995: 148) calls *cardinal virtues*, ‘since from them all others follow practically, and may be derived theoretically’. The fascinating thing about Schopenhauer’s argument is that he does not derive the moral principle and the cardinal virtues from transcendental moral reason but from a faculty well installed in human nature: natural compassion (*natürliches Mitleid*):<sup>21</sup>

It is simply and solely this compassion that is the real basis of all voluntary justice and genuine loving-kindness. Only insofar as an action has sprung from compassion does it have moral value; and every action resulting from any other motives has none. (Schopenhauer, 1995: 144)

As a matter of fact, Schopenhauer uses ‘compassion’ not in the Aristotelian sense, but in the modern sense of *sympathy* – that is, as the capacity to feel

others' feelings as ours, together with the corresponding motivation to alleviate their sorrow or misery.<sup>22</sup> The derivation is innovative because sympathy – as it is generally accepted – is a *partial* emotion which varies according to the (emotional) distance between the sympathizer and the other person,<sup>23</sup> while morality and justice require the impartial point of view of the *law* – that is, 'reason unaffected by desire',<sup>24</sup> as Aristotle would say. Martha Nussbaum states that both standpoints can and must be put to work together in a *complementary* way, although she admits that compassion by itself cannot yield definite normative results and is fallible without an external normative theory (Nussbaum, 2001: 340–342). In fact, she recognizes the superiority of the Rawlsian strategy of grounding justice in a combination of prudential rationality and constraints on information (*veil of ignorance*) over a strategy based on compassion and information (Nussbaum, 2001: 340). Adam Smith's synthesis between sympathy and impartiality through the hypothesis of the impartial and well-informed spectator defines another important strategy for founding moral judgement (Smith, 1984: 294) – that is, to decide upon the *propriety* of conduct: the impartial spectator would identify the proper affection behind prudent, benevolent and just actions and would sympathize with it. As the impartial spectator is not a man without, but the 'man within', he is the key condition for both moral judgement and moral conscience. But if Rawls himself is right, the perfect knowledge, sympathetic feelings and impartiality that characterize the judicious spectator can only render a correct estimate of the net sum of satisfaction – that is, a classical utilitarian justification of social justice – while 'mutual disinterestedness subject to a veil of ignorance leads to the two principles of justice' (Rawls, 1971: 187). It is not only social justice that allows for a utilitarian justification in Smith's theory of moral sentiments. In fact, he also develops an explicit utilitarian conception of universal benevolence (Smith, 1984: 235–237). With respect to Schopenhauer's moral principle, it can certainly be understood in utilitarian terms as soon as we admit trade-offs between the two cardinal virtues (more or less 'justice' in order to reach more or less 'philanthropy'). But the relationship between the two cardinal virtues could also be established as a serial or lexical order, in the Rawlsian sense, with the first virtue prior to the second.

Be that as it may, one interesting thing about Schopenhauer's argument is that the two standpoints (the partial standpoint of compassion and the universal standpoint of morality) neither conflict nor mix, but nonetheless one is *derived* from the other. In addition, the derivation not only gives motivational support to morality, since the moral principle is anchored in emotion, but also yields a *unique* – but compound – general moral principle.

Do the cardinal virtues regulate the emotional economy of cooperation? When somebody cheats you or free-rides on you, as we have already seen, you feel anger and indignation because you come to believe that the other person despises your generosity, offends you and undeservedly takes what does not belong to him. In both cases there is a clear violation of the virtue of justice: *neminem laede*. On the other hand, the first movement of the strong reciprocator, the generous and altruistic decision to help or care, seems to be under the jurisdiction of the virtue of charity: *Omnnes, quantum potes, juva*. Cooperation thus appears as a unique scenario where both cardinal virtues contribute to regulate human behaviour. This seems so clear that we can affirm that any social relation – society in general – could not be said to be moral (that is, regulated by the virtues of justice and charity) unless it is cooperative and mainly guided by rules of reciprocity and reciprocal altruism.

Now we want to finish our argument with a general hypothesis, albeit tentative: if strong reciprocity and reciprocal altruism have co-evolved with our social emotions to afford human cooperative behaviour, as for example Bowles and Gintis (2011) have recently defended, and if cooperation has been the main adaptive advantage in human evolution, then there might be good reasons to think (or to ground the hypothesis) that morality, as expressed in the above general principle, has also co-evolved with the rest of human cognitive-emotional architecture or, at least, is not contradictory to that evolution. It is this evolutionary perspective that gives special importance to Schopenhauer's derivation of morality from the *natural* emotion of compassion.

### Acknowledgements

I would like to thank Francisco Herreros for his insightful comments on a previous version of this paper. I am also indebted to the numerous detailed and relevant comments of two anonymous reviewers from *Rationality & Society*, which forced me to reconsider some topics and refine several arguments. Needless to say, any remaining shortcomings are entirely my own responsibility.

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Notes

1. 'Adfectus quidem tam mali ministri quam duces sunt'. Seneca, *De ira*, I, ix.1.
2. For a brief but well-informed reconstruction of western philosophical views on emotions, see Solomon (2000).



3. For a comprehensive presentation see de Sousa (2003). A very intelligent and open-minded defence of the 'cognitive theory' of emotions is provided in Solomon (2003).
4. The notion of *appraisal* is a key concept in many relevant psychological theories of emotions. See, for example, Lazarus (1991), Frijda (1986), and Scherer et al. (2001).
5. R. Descartes (1649), *On the Passions of the Soul*, The First Article, Indianapolis, IN: Hackett, 1989, quoted by Solomon (2003: 4).
6. To wit: the conception, first, that grounds the trustworthiness of the potentially trusted person in moral commitments and, second, the conception that grounds it in the trusted's psychological or character disposition to be the relevant kind of person. All three are cognitive theories based on the agent's rational or reasonable expectations and prudential judgement.
7. See Aristotle (1994–2008), *Rhetoric*, II, 7 (1385a).
8. As we shall see later, sympathy is not the same as compassion. The latter implies sympathy, but restricts it to cases of *undeserved* pain or disgrace.
9. See Aristotle, *Nicomachean Ethics* (1999: ch. VIII). As is well known, for Aristotle 'perfect friendship' implies *reciprocal benevolence* because it is based on virtue, not on interest as happens with 'accidental friendship'.
10. Aristotle (1994–2008) insightfully sees that fear (*phóbos*) always leaves some room for a certain amount of hope (*elpída*): 'fear sets us thinking what can be done, which of course nobody does when things are hopeless' (*Rhetoric*, II, 1383a). Aristotle analyses the concept of *thárros* (confidence, courage) and *phóbos* (fear) as opposed concepts, and keeps hope (*elpída*) in a close association with *thárros*. This emotional state of the mind (confidence or courage: *thárros*), says Aristotle (1994–2008), 'is the opposite of fear, and what causes it is the opposite of what causes fear; it is, therefore, the expectation associated with a mental picture of the nearness of what keeps us safe' (*Rhetoric*, II, 1383a). Our own contrast between hope and fear is thus compatible with Aristotle's opposition between *thárros* and *phóbos*.
11. Aristotle (1994–2008), *Rhetoric*, II, 2 (1378a), trans. by W. Rhys Roberts.
12. See Note 11.
13. Aristotle (1994–2008), *Rhetoric*, II, 9 (1386b), trans. by W. Rhys Roberts.
14. See Note 13.
15. In this sense, shame is disconnected from guilt. I can feel ashamed but not guilty if I take the vice attribution to be wrong, and I can feel guilty but not ashamed if my vice or my bad behaviour is not discovered.
16. Plato (1970), *The Laws*, I, 647a–649a.
17. See Elster (1989: 111–125).
18. In a negative way: do not do to another what you do not wish to be done to you. In a positive way: do to another what you wish to be done to you. As is well known, Hobbes designates this principle as the second natural law, which can be derived from the first natural law (liberty to preserve oneself). See Hobbes (1985), *Leviathan*, Ch. 14.

19. 'The absence of all egoistic motivation is, therefore, *the criterion of an action of moral worth*' (Schopenhauer, 1995: 140). For a classical critical review of Schopenhauer's moral principle, see F. Nietzsche (1989 [1886], Part V, section 186), *Beyond Good and Evil*.
20. There is some resemblance between this moral principle and Hutcheson's public desires rooted in universal benevolence: 'the general calm desire of the happiness of others, or aversion to their misery upon reflection' (Hutchenson F (2002), Treatise I, Sect. II, pp. 31–32). The difference rests in the philosophical derivation. Likewise, Smith accepts the universal observation of the moral principle but, contrary to Hutcheson, he does not derive it from a 'moral sense' or 'the soft power of humanity' but from the reflective reason of the 'inhabitant of the breast, the man within, the great judge and arbiter of our conduct... the eye of this impartial spectator. It is he who shows us the propriety of generosity and the deformity of injustice; the propriety of resigning the greatest interests of our own, for the yet greatest interests of others, and the deformity of doing the smallest injury to another, in order to obtain the greatest benefit to ourselves' (Smith, 1982: 137).
21. For the philosophical derivation, see Schopenhauer (1995: §17 and §18).
22. Smith (1982: 10), for instance, defines sympathy as 'to denote our fellow-feeling with any passion whatever', but the motivation to alleviate others' sorrow depends on information about propriety and merit, much in line with the Aristotelian argument.
23. See Goldie (2000: 216–217).
24. Aristotle (1984), *The Politics*, III, 16 (1287b).

## References

- Aristotle (1984) *The Politics* (trans. C Lord). Chicago, IL: The University of Chicago Press.
- Aristotle (1994–2008) *Rhetoric* (trans. W Rhys Roberts, The Internet Classic Archive). Available at: <http://classics.mit.edu/Aristotle/rhetoric.html>
- Aristotle (1999) *Nicomachean Ethics* (trans. T Irwin). Indianapolis, IN: Hackett Publishing Company.
- Arrow KJ (1974) *The Limits of Organization*. New York: W. W. Norton & Company.
- Bowles S and Gintis H (2011) *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton, NJ: Princeton University Press.
- Cosmides L and Tooby J (2000) Evolutionary psychology and the emotions. In: Lewis M and Haviland-Jones JM (eds) *Handbook of the Emotions*. New York: Guilford Press, pp.91–115.
- Damasio A (2003) *Looking for Spinoza*. Orlando, FL: Harcourt, Inc.
- DeQuervain D, Fischbacher U, Treyer V, et al. (2004) The neural basis of altruistic punishment. *Science* 305: 1254–1258.
- Descartes R (1989) *On the Passions of the Soul*. Indianapolis, IN: Hackett.
- De Sousa R (1987) *The Rationality of Emotion*. Cambridge: Cambridge University Press.
- De Sousa R (2003) *Emotion*. Stanford, CA: Encyclopedia of Philosophy.

- De Waal F (2006) *Our Inner Ape*. London: Penguin Books.
- Elster J (1989) *The Cement of Society*. Cambridge: Cambridge University Press.
- Elster J (1999) *Alchemies of the Mind*. Cambridge: Cambridge University Press.
- Frijda N (1986) *The Emotions*. Cambridge: Cambridge University Press
- Gintis H (2009) *The Bounds of Reason*. Princeton, NJ; Oxford: Princeton University Press.
- Goldie P (2000) *The Emotions, a Philosophical Exploration*. Oxford: Clarendon Press.
- Hardin R (2006) *Trust*. Cambridge, UK: Polity Press.
- Hobbes T (1985) *Leviathan* (ed. CB MacPherson). London: Penguin Books.
- Humphrey N (2002) *The Inner Eye*. Oxford: Oxford University Press.
- Hutchenson F (2002) In: Garrett A (ed.) *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*. Indianapolis, IN: Liberty Fund.
- Lazarus R (1991) Cognition and motivation in emotion. *American Psychologist* 46: 362–367.
- Nietzsche F (1989 [1886]) In: Kaufmann W (trans.) *Beyond Good and Evil*. New York: Vintage Books.
- Nussbaum MC (2001) *Upheavals of Thought*. Cambridge: Cambridge University Press.
- Pinker S (1997) *How the Mind Works*. London: Penguin Books.
- Plato (1970) *The Laws* (trans. TJ Saunders). London: Penguin Books.
- Rawls J (1971) *A Theory of Justice*. Oxford: Oxford University Press.
- Scheff TJ (2000) Shame and the social bond: a sociological theory. *Sociological Theory* 18(1): 84–99.
- Scherer K, Schorr A and Johnstone T (eds) (2001) *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford: Oxford University Press.
- Schopenhauer A (1995) *On the Basis of Morality* (trans. EFJ Payne). Indianapolis, IN; Cambridge: Hackett Publishing Company.
- Smith A (1982) In: Raphael DD and Macfie AL (eds) *The Theory of Moral Sentiments*. Indianapolis, IN: Liberty Fund.
- Solomon RC (2000) The philosophy of the passions. In: Lewis M and Haviland-Jones JM (eds) *Handbook of Emotions*, ch. 1. New York: The Guilford Press, pp.91–115.
- Solomon RC (2003) Emotions, thoughts and feelings: what is a ‘Cognitive Theory’ of emotions and does it neglect affectivity? In: Hatzimoyis A (ed.) *Philosophy and the Emotions*. Cambridge: Cambridge University Press, pp. 1–18.