

Anton Leist

Potentials of Cooperation

Abstract: Since Hobbes' *Leviathan* was published in 1651, the 'problem of order' has been known for some time. Despite this long gestation period for social theory even today we do not have a universally agreed upon answer to this 'problem'. One of the reasons behind this lacuna may be the overly dispersed work being done in the economic and sociological traditions. Whereas one tradition favours 'collective action' as a central answer, the other thinks of the problem itself being dissolved by the acceptance of 'socialized man'. Here, an attempt is made to offer the phenomenon of 'cooperation' as a promising middle ground for both traditions. To underline the importance of cooperation as an elementary social activity, first, cooperation is shown as working in tandem with its rival 'competition'. Secondly, several conceptual analyses of what is included in collective action and cooperation are offered. These analyses, thirdly, are deepened by an overview of the motivational bases potentially advancing cooperation. Overall, an awareness of the self-creating character of cooperation is explored, and put forward as the most feasible way of answering the classical problem of order.

1. The Problem of Order

Quite a bit of philosophical curiosity seems to be needed if, in approaching an understanding of social matters, one throws up the one 'first' question: how is social order in society possible? This question has a similar ring to the definitely metaphysical question: 'why is there anything and not nothing?' The order-question, famously put at the beginning of social theory by Talcott Parsons (Parsons 1937), angles towards philosophical nosiness, only with a more narrow focus: 'why is there order in societies, and not disorder?' Admittedly, the question of social order is, at least since the arrival of the classic sociologists documented in Parsons' treatise, more on the mind of social scientists in search of a 'global theory' rather than being a pet subject of philosophers. But for these scientists, too, the handicap lies in the question's being too global to allow one elementary answer that is fitting for heterogeneous societies. So far as answers have been given, they precariously drift into trivialities or become labels for patchwork collections of micro-explanations. This might also be true for one, at first sight rather suggestive because utterly simple, answer: social order exists in a society because of *cooperation* among its members.

There are at least three kinds of reasons why the ‘cooperation’-answer again might share the fate of other proposals. A first reason, perhaps the most obvious, lies in the heterogeneity of people’s social motivation. People seem quite uniform in their self-interested motivation in the sense that everyone has it in some degree. Otherwise they would be incapable of leading a life of their own. People are pretty diverse, however, in the extent they relate to other people, contribute to common goals, participate in social movements, etc. We could call this first aspect the ‘*social-motivation heterogeneity*’.

Secondly, and perhaps less obviously, there are different ways of ‘relating’ to other people, given a motivational disposition. Structurally individual and social action do not simply fall into two separate homogeneous classes, and however one analyses individual action, social action will be more complex, simply because at least two agents and often many more are involved. For example: individual agents can, roughly, act individually, but with an awareness of other individuals’ behaviour, they can act together explicitly on the basis of individual aims and intentions, and can explicitly and openly share collective goals with others. These are three kinds of social action, which can be analysed in even finer detail. Let us call this second aspect the ‘*social-relation heterogeneity*’.

Thirdly, the world of a society is not only peopled by individual agents, but also by institutions and collectives of different sorts. Individual agents are involved in these collectives in various ways, representing, promoting, controlling or being in general (and often unconsciously) normatively tied to them. This is especially relevant for collectives and institutions of education, production and politics. These institutions are varied in size and kind, which makes for different ways of individual agents being related to them. The most striking difference may be the one between free associations such as firms and universities, and the state actively imposed on citizens through its institutions. Let us call this third aspect the ‘*institution-heterogeneity*’. This third category will become still more complex if general social norms (besides those of institutions) are added to it.

Interest in cooperation, then, can be rephrased like this: given these dimensions of structural heterogeneity in motives, social relations and institutions, which role does cooperation play among them, and which sense of ‘cooperation’? Why think of cooperation as paramount among other forms of social connectedness in answering the problem of order? The basic idea could be this, beginning with motivations. However we think of the motivations people have towards other people and towards society, it seems that some kind of coordination of single actions is needed in order to achieve collective goals, thereby improving the living conditions for all. Admittedly, coordinated social action without a collective goal can be advantageous in itself, as becomes obvious from individual driving according to commonly accepted traffic rules. But if all of society’s social action would be similar to road traffic, there would be nothing to drive with or to drive on. The three social spheres mentioned—education, production and politics—need collective action in a more deeply integrated way: they need social action with collective goals. And it seems that, unavoidably, there is no other name for this kind of action than ‘cooperation’.

Does cooperation, on this basic level, have any socio-ontological rivals? *Competition* and *social control* spring easily to mind, the latter in an extreme form as *suppression*; not so readily, perhaps, one thinks of a functionalist view of parts of society ‘larger’ than the individual. Competition and control are phenomena at the level of actions and agents; functionalism is at the level of ‘social systems’ (including systems of norms). How do these three types of social connectedness (competition, control, function) stand with regards to cooperation? We can, against functionalism, defend the principle of ‘methodological individualism’. The basic idea of such a defence is that supra-individual entities are theoretically dispensable if individual entities are sufficiently explanative.¹ That leaves the two alternatives to cooperation, competition and control, at the level of individual actions and agents. What should we make of them?

If there is, with the move just mentioned, a *first priority* of individual action against the potential action of collectives and other supra-individual entities, there also is a *second priority* of free (individual) action against controlled or suppressed action. The argument to think of is involved in a Kant-like thought experiment of universalization. If everyone controlled everyone, this would be a breakdown of the opportunity of controlling. In Kant’s famous example, controlling or suppressing can no more be universalized than lying. This thought experiment establishes the *normative* (even if, alas, only normative) priority of free action to a suppressed one, and excludes the large-scale dominance of ‘power’. Society cannot be built basically on universal social control or suppression, even if compatible with quite an amount of it empirically.²

Presupposed by these two priority arguments is a still more fundamental, *zero-priority*-argument, which I will not go into more explicitly: the priority of *rationality* against irrationality. (Some call this the ‘principle of methodological rationalism’.) The clearest argument for it comes from Donald Davidson, through the idea that intersubjective understanding presupposes rationality in a minimalist sense, and that ‘rational’ action as means-ends-action is therefore of higher priority than irrational action. Again, this is a normative priority that allows an open-ended but not encompassing amount of irrational action.³

This leaves us with the last alternative in the individual-action branch: *competition*. Is there another priority-argument to the benefit of cooperation as normatively prior to competition? Even if, taking clues from the public goods problem, a priority of cooperation over competition is often taken for granted, given the generality of the question the evidence for the claim seems a bit overstressed.⁴ How important are public goods for society in general, and is it true

¹ For defence and clarification see Elster 1982; 1985, ch. 1; 1989, 105–6; ch. 3. The principle should not be identified with the narrower principle of ‘methodological egoism’, a principle strongly connected with *homo oeconomicus*. I will come back to this in the next section.

² One way to evade this conclusion would be, though, making meaning changes in the concepts of control or power, as is being done in part of the literature. ‘Control’ could mean, for example, only supervision without sanctioning.

³ See Davidson 2001, especially “Rational Animals”. MacDonald/Pettit 1981 pointed out Davidson’s relevance for social methodologies quite some time ago.

⁴ That would be true also if one tried to argue from within a conceptual analysis of ‘cooperation’ and tried to make room for competition on a sublevel of coordinated action, described overall as cooperation. See, for example, Michael Bratman’s attempt to locate chess playing

that solutions cannot be reached through competition? In order to work out a more detailed argument, I will, in *section 2*, make use of an exchange between James Buchanan and David Gauthier on moral preconditions for an ideally free and efficient market. Gauthier developed his view through a defence of a normatively ‘Lockean proviso’, something suitable to show the necessity of cooperation for even something as ideal as the ideal market.⁵

The focus on the phenomenon of cooperation thus opened, *section 3* will be devoted to the ‘conceptual’ side of it, proceeding with an eye on potential explanations rather than in order to logically restrict empirical research. *Section 4* will develop an overview on different classes of potential motives for cooperation. The most important among such motives are those of self-interest or rationality (4.1), identification with collectives (4.2), ‘altruist’ motives (4.3) and motives provided by social norms (4.4). Lastly, in *section 5*, I will point out why cooperation—on the back of these various diagnoses of potentials for cooperation and also in the face of some degree of pessimism—nevertheless is comparably the best conceptual focus for bringing us closer to answering the problem of order.

2. Cooperation and Competition on the Grand Scale

2.1 Preplay

A classical argument for the theoretical significance of cooperation runs like this.⁶ Given the perfect market paradigm of rational, unconstrained and non-cooperative interaction, there would be no occasion for cooperative or collective action if there were not ‘market failures’. Market failures bring collectively suboptimal outcomes. Outcomes are suboptimal when it would be possible to make some individuals better off without making others worse off. Defenders of the perfectly competitive market (PCM) grant the practical relevance of morality, law and politics because of market failures in this sense, with different emphasis among themselves to the extent and necessity of these failures. This concession to morality, law and politics can be extended to cooperative action, as cooperation is also dependent on (or strongly linked with) morality, law and politics.⁷ The unavoidability of cooperation, then, is understood, even if only on the basis of a PCM and its conceptual elements. What are these elements?

PCM comes along with, especially, a view of individuals as self-interested, perfectly rational, uninterested in others, unrestricted by norms. Consistent with widespread use, in the following I will always call such action or agents ‘rational’, contrasting them with other normative forms of agency as ‘reason-

in an analysis of cooperative action: Bratman 1992, 332. Such an argument would be constrained by being only conceptual, leaving out the role of motivations. This constraint will become important in my further argument.

⁵ In following Gauthier’s argument I accept the necessity of idealizations. And I bypass the potentially misleading claim made by Gauthier that the market is a ‘morally free zone’. As will be shown, Gauthier himself does not stick to this claim as strictly as it may seem at first.

⁶ The thread I am following here comes from Coleman 1987 who is referring strongly to Knight 1951. I will also make use of Gauthier 1986, especially chs. 4 and 7.

⁷ I will make some short remarks on these dependencies in *section 5*.

able'.⁸ Bearing the contrast between cooperation and competition in mind, the difference between individual rationality and social rationality of some kind—emotional or moral in the latter case—strikes one as most relevant. To accentuate the contrast as simply as possible: if there are, ideally, rational agents with *individual* goals and ones with *social* goals ('social' including at least two agents), why start in social theory with the former and not the latter, or with both in some combination?

Three prominent reasons (more or less explicitly normative) for the priority of the rational individual could be called the '*anthropological*', the '*universalizability*' and the '*optimization*' reason. The anthropological reason, be it backed empirically or not, identifies individual self-interest with claims to 'human nature'. Plausible as this idea may have been in the 18th century, in the face of extended knowledge of non-Western societies it is difficult to rely any longer on a substantial version of individual rationality as part of 'human nature' (for relevant ethnological research see, for example, Hann 2006).

The second reason applies the universalization test both to egoism and altruism. To some, a fully egoistic society seems possible, a fully altruist one not (see Elster 1989, 36). The more explicit argument for the latter statement seems to run as follows: if all are altruists and put the interests of others ahead of their own, this would run into a self-contradiction or regress, as all had to take their own interests into account, which are the most important interests of all altruists on the primary level. But this seems informative only for highly artificial altruists, whereas the everyday altruist must keep looking for at least some of his own interests in order to be capable of helping others. The normative priority shown by the asymmetry seems not to be one relevant for everyday altruists.

The third reason—the one from optimization—connects the argument with a specific theoretical view, namely with PCM. This optimization argument states that systems of social exchange populated with rational individuals are the most efficient ones among all available alternatives. Normally efficiency is measured in terms of individual material gain. But it could also be framed more abstractly, for example in terms of social stability (something suggested by Knight 1951). The suggestion is that a market based on individual rationality would be more stable compared to social systems realized through value agreements. One can be sceptical regarding this point, but the argument is more complex, as it depends on how ideal the market is taken to be. Stability of a social structure is useful as a criterion only in near-to-reality conditions. Stability within PCM would be analytic and could easily be opposed by analogously analytically stable communities based ideally on homogenous social identities. Even if 'stable' may be a relevant criterion for efficiency, its testimony is dependent on the level of idealization, even if idealization can hardly be avoided.

⁸ The terminology comes from John Rawls: see, for example, Rawls 1993, 48–54. I will use 'self-interested' and 'egoistic' interchangeably, even if there are different connotations and contents attached. Also, catching self-interest with 'rational' is not strictly adequate, as one can be rationally utility-maximizing in regarding others. See also the difference between methodological rationality and methodological egoism mentioned in *section 1*. 'Rational' in the following is meant as an abbreviation, to be completed by context.

Let us be clear, therefore, about the state of the dispute springing from the third reason. Defenders of the priority of competitive action accept the importance of cooperation, but only as a corrective for pockets of failure in the overall successful competitive scenario, itself based theoretically on PCM. According to PCM, competitive rationality and the rational agent are ‘basic’ for society; cooperation and its normative elements, morality, law and politics, come into play only secondly because of the underdeveloped form of PCM in an actual society. Ideally or ‘in principle’, a society could be realized fully according to PCM. Given this defence of a priority claim for competitive action, there are two ways to object.

A first would be to challenge the idealization claim. One could question that PCM is realizable and that its ideal description, internally coherent as it may be, is normatively relevant under the conditions of actual societies. If this objection goes through, however, it also removes the opportunity to diagnose the relative importance of competition and cooperation to each other. General statements on their mutual relation need idealizations, but the question of what is left ideally if PCM is challenged remains open. Instead of objecting *in toto* to the idealization, then, one might better inquire into a gradual de-idealization of PCM. Is the ideal market indeed, as PCM claims, not itself in need of some form of cooperation? Is indeed only market failure under nasty real conditions ushering in cooperation second-handedly?

At this point the debate shifts into one concentrating on the importance of normative presuppositions for PCM, especially those of a moral and legal kind. And the best way to put the claim *for or against the priority* of competitive action is whether or not such normative conditions for the market themselves can be based on rational action alone. This seems necessary, because otherwise the defence for the priority of competitive action would trivially break down. All defenders of PCM accept that even an ideal market can only work under social conditions such as stable property rights and the absence of force and fraud. Competitors in the market have to accept a normatively constrained form of competition, be it through moral dispositions or, as libertarians prefer to argue, institutions of law, again perhaps based on rationality. How, now, to decide between the moral/legal or the rational as basic?⁹

2.2 The Buchanan-Gauthier Exchange

All defenders of PCM accept that there is, in a global sense, an ‘ethic’ to be presupposed and with it a certain social bond. But according to them this ethic seems to rest on a narrow basis of rationality and freedom, and not on something mysterious and contested like moral dispositions. The claim on the side of PCM-defenders is, then, as follows: the ideal market is ‘morally neutral’ though collective goods like the security of rights are involved. It is built on rational motives alone and the distribution within ideal market-exchanges is rationally fully satisfying because a fully proportional outcome for the inputs

⁹ In the following I will forego the more complex level of the legal as it seems to rest, at least partly, on the moral.

of every participant will be realized. Neither need there be recourse to moral motivations, nor would there be even the logical occasion (within the logical construct of PCM) for moral motives. What is ideally rational does not need morality.

Some light can be shed on this argument by recalling the exchange between James Buchanan and David Gauthier on the latter's explanation of the market as morally constrained by impartiality, but largely based on rationality.¹⁰ It is interesting to see in this exchange, first that de-idealisation of PCM needs some conceptualization of the starting conditions for economic exchange, and secondly that different opinions on these conditions can only be defended on the basis of empirical hypotheses, especially ones concerning social stability. Even if crucial statements at this point of the story typically make the empirical dependent also on the effect of rationality-claims, there is always a reference to actual behaviour involved. Let us see how this is set to work.

Buchanan considers himself to be a true defender of the market as a purely rationality-driven sphere, as he declines the use of morality in any sense in the turnover from a Hobbesian 'state of nature' to the market. Buchanan offers a Hobbes-like transition-scenario from the 'natural equilibrium' distribution to a 'constitutional contract'. Yet contrary to Hobbes Buchanan thinks that the state of nature is not full anarchy, but involves a natural equilibrium in the form of a stabilized condition among all rational individuals—stabilized especially through an effective match of predatory and defensive expenditures. This situation, even if stable, is suboptimal as the costs both of predation and defence are unproductive. For Buchanan the difference between the suboptimal stage and a rationally improved stage—even one conserving the natural distribution of power—is enough to guarantee a cooperative outcome. "I commence from the *status quo* distribution of rights and I do not apply criteria of justice to this distribution." (Buchanan 1988, 85)

Gauthier, on the other hand, thinks this transition to be irrational, as it would "invite others to engage in predatory and coercive activities as a prelude to bargaining. [...] Co-operative compliance is not compliant victimization." (Gauthier 1986, 195) If a society of masters and slaves is to be expected to turn into a society with free exchange, the effects of earlier repression must be eliminated. The conditions for free exchange suggested by Gauthier are following the idea of a *coincidence of rationality and fairness* (as impartiality), the latter being fixed technically by a specific version of the 'Lockean proviso' (Gauthier 1986, ch. 7). Gauthier thinks that rational slaves would not comply with asymmetric roles bearing the mark of former coercive interaction. Rational masters, on the opposite site, would accept the rational advantages of free cooperative production and fairness in the sense of not worsening the position of others within potentially cooperative relationships. *Rationality* as improving one's utility and *fairness* through reciprocal exchange would fall into one.

¹⁰ See Buchanan 1988 and Gauthier 1986, esp. ch. 4. Buchanan argued on the basis of his 1975.

One might be sceptical towards this conclusion.¹¹ As not least Buchanan's reaction makes clear, there are divergent psychological presuppositions in play. And indeed, rationality here is not to be understood as logical coherence but as depending on psychological dispositions in actual social contexts. Whether and how the transition from a state-of-nature condition into constrained maximization would be successful is open to empirical judgement. It is not detrimental to his claims for rationality, therefore, if Gauthier backs up his argument like this:

“Co-operation on terms less than fair is therefore less *stable*, in failing to gain the whole-hearted acceptance of all participants. Fair co-operation invites a full compliance which each does not stand ready to withdraw because of shifts in the natural distribution. For although no one is prepared to concede advantage to others in order to bring about this *stability*, each is willing to accept it, as enhancing the benefits of co-operative institutions and practices, when it is not costly to herself to do so.” (Gauthier 1986, 230, emphasis added)

Where does this leave us with the question we started from in this section, the priority of either competition or cooperation? If we accept Gauthier's line of argument, the conclusion would be that rationality *and* morality, utility maximization *and* fairness in combination are a precondition of the market and competitive behaviour.¹² It would be wrong to simply claim that ‘the market is based on morality’, as rationality and morality seem to be intertwined too narrowly to make such a blunt claim informative.¹³ Being interested in the role of cooperation, there has to be cooperation presupposed in the sense of keeping to a sense of fairness, which means as much as internalizing potentially negative costs for others. For a market to be effective what has to be involved is not only recognizing the liberty of each, but also fair behaviour in not reaping one's benefits by imposing costs on others. If this is meant, as Gauthier suggests, not only for actual and immediate contributions but also for social structures in general, fairness objects to social conditions restricting the freedom of some (say, of women) to the advantage of others (say, of men). We can leave it open here as to which extent this extended demand of fairness needs a welfare state to accompany an effective market.

The conclusion, if, in this series of arguments, we stick more closely to Gauthier's opinions, is that cooperation seems not to be a solid block on its own on which to build a competitive society (something anyhow quite implausible), but rather a quality of the very relationships that are at the same time rationally

¹¹ The Hobbesian lingo to express this scepticism is provided by Buchanan: “Recognizing the prospect that the slave might not rationally comply with an agreement, the master, before agreeing to terms, can communicate to the slave that any departure from the terms will bring punishment. And, indeed, the agreement itself may include the establishment of an effective enforcement agency.” (Buchanan 1988, 85)

¹² For stressing these combined conditions see Gauthier 1986, 262 *passim*. See also his clear statement as to cooperation: “Society, as the framework within which market competition and co-operation occur, is itself a single co-operative enterprise.” (Gauthier 1986, 260)

¹³ Such claims, however, have often been made, for example following Weber's thesis of protestantism. See, among others, Koslowski 1982.

effective. Thus cooperation, and with it the sources of morality, law and politics, is not something secondary in rationality failures, but a quality at the basis of all efficient social relationships. Can this insight, abstract as it is, now be made more concrete?

3. Logical Forms of Cooperation

To take up this task is all the more important as so far the argument has said nothing about the possible *reality* of a fair market, be it one according to Gauthier's Lockean principle or one with different (or differently applied) principles of fairness. To make good on this lacuna, think of two forms of heterogeneity mentioned at the beginning: those of motivation and those of relation. Different motivations and different ways of being related to others show up in different forms of cooperation. People come to cooperation with different motives, and expect different kinds of cooperation. In order to develop a realistic sense of *potentials* for cooperation, these alternatives have to be brought into the open.

It would be wrong, however, to think of motivation and relation, the latter as forms of cooperation, as strictly distinct. Such a perspective would assume people are equipped with motivations *independent of* social relationships but are ever ready to enter into them. According to this picture people instrumentally choose relations fitting their independently given motives. Everyday experience already shows this picture to be false. There are interdependencies between motivations and forms of relation. To put it rather starkly: available relationships and the behaviour of others *constitute* a motivation for the agents involved, out of given dispositions for their motives being actualized in one way or another.¹⁴ Different social contexts, then, provide the scene for agents to act differently within them. According to this *priority of social context* over and against isolated agents we should first look at contexts, and only later at individual action and its logic.

'Social contexts' come in many degrees and forms, however. At the elementary level addressed here it makes sense to distinguish between 'contexts' in form of the 'logical structure' of what cooperation could mean, and the effective behaviour towards and of others as studied by empirical researchers. To disentangle *possible ways of cooperation* and *actual motivated behaviour* is only possible at all because the specific forms of 'collective action' and 'cooperation' can be analysed analytically by referring to the agents' intentions and beliefs, but ignoring their motivations. Indeed, this is how analytic philosophers typically proceed in their analyses of social actions.¹⁵ Even if we do not know much about the empirical relevance of the logical forms of collective action after reading their analyses, to know about them seems to be a relevant first step. Accordingly, I will make some distinctions as to the forms of collective or cooperative action

¹⁴ This rather strong statement will amply be underwritten by glosses on recent empirical research of behaviour in situations with conflicting material interests, to be covered later in *section 4*. See especially the end of *section 4*.

¹⁵ I am referring here to the work of Michael Bratman, Margaret Gilbert and Raimo Tuomela. All of their analyses are untainted by references to motivation.

in this section, leaving the next section for an overview of differently motivated cooperation.

First we must take note of the simultaneous use of two terms, ‘collective action’ and ‘cooperation’. ‘Collective action’, because of being a wider term than ‘cooperation’, is also a more complex one. In the following I will not refer to the action of collectives, but only to individually collective action.¹⁶ Even then, due to different uses in economics and other social sciences, ‘collective action’ and ‘cooperation’ relate to each other differently. Let us first concentrate on ‘cooperation’. In order to find out what we mean by this, we can look at the possible range of its use, from the narrowest to the widest versions of it.

In such a process, it seems easier to think of the narrowest version rather than the widest, the distinction between ‘narrow’ and ‘wide’ signalling here a stronger vs. weaker degree of interconnectedness between agents. The narrowest version is perhaps best illustrated by an example such as playing a sonata together. People engaged in such an activity have a common aim, the performance of the sonata, and they contribute to the aim according to a plan of contribution, taking note of each other’s actions in the process. In the full sphere of social action this kind of cooperation is quite rare, as it requires an extremely high level of mutual control and responsiveness in order to achieve the common goal, the performance of the sonata. Most other forms of cooperation, and especially those in common work processes, do not need this degree of tight mutual control and response.

(C1) People *cooperate* if and only if they know of, and intentionally contribute to, a *common goal*, by intentionally coordinating their individual actions in a way presumptively conducive to the realization of the common goal.

The crucial content of what cooperative action asks for here is in the formulation to intentionally coordinate “their individual action in a way presumptively conducive”. Following Bratman (1992) this could be explained as including two elements: a preparedness mutually to *respond to* (i) and *support* the others in their efforts (ii). In the sonata example, normally responsiveness will suffice, as perhaps a cello will not be able to adopt the violinist’s part. In most examples of cooperation contributory tasks are not as tightly assigned as they are in making music together, and then mutual support has some scope of application. In general, mutual responsiveness and mutual support are the two necessary conditions in narrow cooperation. In the following I will call this ‘*full cooperation*’. Further aspects in full cooperation, also detectable in other forms of cooperation, will be mentioned shortly.

What do we find at the opposite end of the spectrum? In principle there are two radical alternatives to full cooperation: conflict and indifference. ‘*Conflict*’ would signal action aimed at disadvantaging the other party in some way. ‘*Indifference*’ would represent individual action involving awareness of the existence

¹⁶ I will also make incidental references only to corporate action of individuals. This is action that presupposes collectives like a firm or social institutions, then being represented in individual action by members. Overall the third kind of heterogeneity mentioned at the beginning, institution-heterogeneity, will not be dealt with thoroughly here.

of other agents, but taking them into account only to the extent of favouring or obstructing its own aims. ‘Social’ action is taken here in the Weberian sense of being somehow consciously related to another agent, which makes both types of individual action also social. Indifference would be, for example, the (social) attitude of car-drivers considering whether or when to pass another car, or pedestrians on a crowded pavement trying not to bump into other pedestrians, but otherwise being disinterested in them. There is some ‘regarding’ involved in these cases, i.e. of the other cars or pedestrians, but no regarding under a common aim. It would seem artificial if one imputed an interest in collective safe driving to the driver or an interest in collective unhampered strolling to the pedestrian.

Maintaining the focus on the opposite side of full cooperation, what ought we think of ‘*collective action*’, given that ‘collective action’ seems to be the wider and less socially ambitious concept? What would be the widest form of collective action, now distinct to conflict and indifference? There is only a fine line of demarcation between indifference and collective action in the widest sense marked by a *collective aim* or good. A collective aim would be not to litter in the streets or to drive within speed limits, the individual agents’ motive being directed at least partially to this collective aim. Collective action, distinct from conflict and indifference, could be defined minimally by individual awareness of and some preparedness to contribute to a collective aim.

(C2) People act *collectively* if and only if they know of and intentionally contribute to a collective aim.

To make the contrast clear between C1 and C2, the phrase ‘common goal’ seems a better fit for C1, as there is more intensive coordination involved in cooperation. Minimally a shared plan to achieve the goal is necessary, with some free space for subplans to coordinate plans effectively. Individual actions, plans and subplans have, in a sense, to ‘mesh’ (Bratman) to make the action fully cooperative. Nothing like this meshing is necessary for following a ‘collective aim’ such as not having litter in the street or celebrating a national holiday, something that can be done in extreme individual isolation. What is necessary is only the individual intention to contribute to the collective aim.

Nevertheless, it makes sense to accept different levels of fulfilling C2, taking into account different degrees of interconnectedness of individual intentions. It is here that the notion of a ‘*shared intention*’ to contribute to a collective aim becomes relevant. This notion is at the centre of all conceptual analyses of collective and cooperative action. Whether, and exactly where, to make a distinction between collective and (narrower) cooperative action depends on the degree of interconnectedness of individual intentions conditional for cooperation. There is some space for interpretation here, and different distinctions can be made. But the following seems to me a plausible way of drawing lines between different types of collective action, with ‘collective action’ now used as an umbrella term.

Collective action (shared common goal)			
Non-cooperative collective action: Shared common goal		Cooperation: Shared common goal <i>and</i> shared execution intention	
Individually followed unconditionally	Individually followed conditionally	Full cooperation: One master plan and many/ all subplans shared	Competitive or contingent cooperation: One master plan, but no subplans shared
(Kantian non-litterer)	(reciprocal non-litterer)	(working together)	(bargaining, sports, games, market)

Figure 1: Forms of collective action

To recap, collective action is but one form of social action; conflict and indifference are two others, not directed at a common goal. Collective action is oriented at a common goal.¹⁷ What distinguishes cooperative from non-cooperative social action are shared/non-shared intentions regarding the execution of a common goal. To make such an explicit distinction is *not* usual procedure in the economic literature. Social scientists and philosophers in the economic tradition normally talk of cooperation as purely outcome-oriented, completely leaving aside the distinctions crucial in figure 1 based on degrees of *intention*.¹⁸

This conceptual idiosyncrasy typical for the economic literature explains itself from the economists' behaviour-bound methodology. We should not let ourselves be distracted, however, from taking note of the 'normal', i.e. an 'everyday' understanding of 'collective' or 'cooperative' action, the neglect of which may be justified only within certain forms of research. Consequently one must constantly remind oneself that among social scientists there are regularly two different terminologies in use: 'collective action', in the economists' sense, covers the whole of the types of action above and is treated as extensively equal with 'cooperative action'. A Kantian non-litterer who does not himself litter but is not inclined to take steps towards changing others' behaviour in littering would have to be called, by economists, 'cooperative'. Economists leave it up to empirical inquiry as to which extent an intentionally shared common goal or shared execution plans and subplans are relevant. This should not be misunderstood in the sense that, for them, these ingredients are definitionally *excluded*!¹⁹

¹⁷ In this overview I make use of Bratman 1992 and 1993. Gilbert and Tuomela offer similar analyses of collective action, all through more or less elaborate analyses of 'shared', 'joint' or 'collective' intention. See Gilbert 2008 and Tuomela 2007.

¹⁸ See, as not untypical, Elster 1985, 137; Elster 1989, 24–27.

¹⁹ There are also two different definitional strategies in use. Philosophers define with necessary and sufficient conditions, trying to grasp something like the self-standing entity. Economists define with an eye to the solution of behaviour problems, to be solved on average. They define 'collective action' through the 'problem' of collective action, and which necessary/sufficient conditions are relevant for this to them is an empirical and not a definitional

The economists' definition of collective/cooperative action could be roughly summarized as follows:

(C3) People act *collectively* or *cooperatively* if and only if they contribute positively to an outcome, which is beneficial to (nearly) all and is better than outcomes without contribution.

The more technical definitions in economists' parlance often make use of 'collective' or 'public' goods, which are goods the existence of which is better for (nearly) all, but which involve the problem that non-contributory individual use of the good cannot be excluded and is preferable for rational, egoistic individuals. This leads to the usual understanding of collective/cooperative action on the back of the public-goods problem:

(C4) People act *collectively* or *cooperatively* if and only if they provide outcome-contributions to a public good.

The definitional strategy behind C4 contrasts strongly with the strategy and understanding behind C1, something even more contrastive if one takes the depths of philosophers' explorations of interlocking individual intentions into account. Nevertheless, why should these two strategies not meet each other? Professionally philosophers are in danger of inventing entities the reality of which is difficult to judge. Contact with empirical knowledge might curb the extent of otherwise light-hearted inventiveness. For the economists the contrast between C1 and C4 is a reminder that the object of their investigation is not the real thing. Their light-heartedness towards neglecting the effects of joint-intentions will be documented in the next section. Even if there is not the postulate of a substantial 'we' involved in collective action, some extent of 'meshing' seems necessary.²⁰ It remains to be seen now what this means on the motivational side of cooperation.

4. Four Sources of Motivation for Cooperation

If we want to reach an overview on the auspices of cooperation from the side of motivational potential, and to take this in a context-free general sense to begin with, one clear division impresses itself on us from the remarks in the last section: motivation as underlying the public goods problem, i.e. *rational motivation*, on the one hand, and *residual forms* of motivation on the other. Let us remind ourselves again that this sorting task is not to be understood as

matter. In general, the economists' methodology is the more adequate because of being the more holistic one.

²⁰ Philosophical work on collective action focuses strongly on the 'ontology' of actions and their ingredients. It often asks what has to be postulated in the world for such action to exist. From this background there are also different methodological ideals in play than in the economists' view on action. For an 'individualist' analysis of joint-intentions see Bratman 1993. For clarification of the methodological aspects involved, from the philosophers' point of view, see Gilbert 1989, chs. 1–2, 7. There is little reflection that crosses the gap to motivation, however.

merely distinguishing different kinds of motivation in order to point out different solutions for one and the same problem, cooperation. As has already been said, motive and relation are not independent of each other. Different kinds of motivation come with different perspectives on how to think of cooperation in the first place. This is quite obvious in the approach from rational motivation, but it also holds true for the alternatives. Nevertheless, given the dominance of the rational action approach, it seems best to proceed first with the rational agent in place, and then inquire as to how it can be improved upon or complemented.

To unpack (again) what ‘rational’ action means let us be clear that it includes three elements in particular: the rational in the sense of optimizing ends (i), egoism or self-interest (ii), and ‘outcome-orientation’ (iii). Jon Elster helpfully contrasted the last element with ‘process-orientation’, which profits through participation in acting with others (Elster 1985, 145; 1989, 36–49). Normally, this will not be an either-or-affair, as people are largely engaged in cooperation for outcomes and rarely for mere participation. Pure process-orientation, whether egoistic or non-egoistic, is not only rare but also futile or even dangerous. Elster points to political activism with its highly motivated short-term enthusiasm but overall self-defeating character (Elster 1985, 148). Participation can be of increasing importance, however, once cooperation for outcome is under way.

Somehow the ‘residual’ types of motivation can only be partly of a participatory kind and have to be made up by various types of non-self-interested motives. Complexity arises here, as these motives can be analysed differently, either according to *moral or ethical* categories, or according to *social*, with ‘social’ here being meant as morally/ethically neutral. Typical examples of the latter kind refer to communal identities or to social norms; those of the former include attitudes like altruism or fairness. This distinction between social and moral extensions of the self-interested individual brings up a crucial question: should we think of these extensions primarily as social (and partly moral) or as moral (and secondarily also social) ones? Often argument against the self-interested individual is mixed with, if not based on, moral intuitions, but such an opposition would unfavourably narrow the field.

In order to see the differences, let us define a ‘communal interest’ as an interest in the good of a community, neutral towards the community’s aims. As these aims could, for example, be connected with aggressive nationalism, highly amoral, communal interest is not *per se* a moral motive. To which extent, on the other hand, a moral attitude is also a social or collective one is open to interpretation. The Kantian agent must be a social one, as the Categorical Imperative makes reference to other agents. But does it include joint action? It seems not, because it requires (at least in the standard version) unconditional obedience, irrespective of others’ compliance. The agent has to obey her duties whatever others are doing, which makes her acting a strongly individual, even if in her mind social affair. To sum up, communal social ‘identity’ does not fall into one with a moral ‘identity’, and moral acting at least to one model must not be deeply social.

The same split can be involved in ‘social norms’ compared with ‘moral norms’. Social norms are often defined as sanctioned social expectations, and normally they are based on emotional dispositions.²¹ In order for sanctions to work, social norms have to be shared by a certain number of people, approving or disapproving among each other on relevant behaviour. It is much more difficult to find agreement, unfortunately, on what moral norms are. In the present context it may be sufficient to consider moral norms as directed at the well-being of humans (and partly animals), and social norms can be neutral or even unfavourable towards such a goal. Social norms of dress, for example, can strongly express hierarchical social relationships. Norms of punctuality or orderliness can be instrumental for highly amoral ends. Social norms of help, even, may be criticized from the point of view of morality. Moral norms are often seen as norms of impartial judgement and not as springing from personal attachments, more prevalent in everyday life.²²

As moral identities and moral norms are much more a theoretical, ethical affair with unclear entrenchment in social behaviour, residual motivation to self-interest should primarily be considered as a *social* one, with ‘social’ covering several alternatives. This leaves us, at the first level of opposition to self-interested behaviour (i), with communal interest (ii), altruism (iii) and social norms (iv). *Moral* motives and *moral* norms can then kick in at a second level, perhaps correcting, restricting or changing motivations on the first level.

Motivation	Mode of cooperation	Involvement in action
Self-interest	Conventions Assurance game Prisoners’ dilemma	Outcome orientation
Communal interest, we-goals	Identification with collectives and communities	Process orientation prior to outcome
Altruism	Free-rider punishment in public good situations	Outcome orientation
Social norms	Sanctions Internalization	Heterogeneous views on process and outcome orientation

Figure 2: Forms of cooperation due to different motives

²¹ Definitions of social norms are contested and depend, as is to be expected from the foregoing, on the different traditions of social theory. See more on social norms below, *section 4.4*.

²² What makes the definition of ‘moral norms’ contested is the degree of theoretical perspective controversial among ethicists. There are different opinions as to which extent moral norms have to be social norms at all, i.e. to which extent they must be socially embedded and followed among real people. Elster (1985, 148–151; 1989, ch. 5) helpfully discusses the distinctions between Kantian and utilitarian morality on one side, and altruism on the other. See also Schaber 2011 for the distinction between altruism and morality.

In the following I will give a short overview of the auspices for cooperation under each one of these four categories of motivation: self-interest, communal interests, altruism and social norms.²³

4.1 Cooperation under Self-interest

Analyses for the potential of cooperation among self-interested people have been provided for three types of constellations in particular: one answers a *problem of information* on the basis of non-conflicting interests, a second a problem both of information *and* partially conflicting interests, a third the problem of *conflicting interests* only. In game-theoretical terms these are the problems of *convention*, of *assurance* and of solving the *prisoners' dilemma*. I can here give only a short sketch of these three complexes that have received an enormous amount of attention in relevant literature. Following Elster (1989, 3) it is interesting to note that the 'problem of information' looms large besides the more well-known conflict of interests basic to the 'problem of order'. But it is the second 'problem of order', the one of sufficient information, which corresponds to and partly deepens the classical problem of order.

'Conventions' in the rational-choice perspective are understood as norms for the coordination of actions that are to the benefit of all involved. The classic example for this is whether to drive on the right or the left side of the road. Given a narrower definition like C2, one could not call collectively driving on one side cooperation, as there are hardly joint intentions involved, and not at all meshing execution plans. There may be joint intentions and also plans at times of installing a convention. Once the convention is in use, however, behaviour develops to what I have termed earlier 'indifference'. It loses its collective character, as it develops into social behaviour without joint intentions.

What is of interest is less how to label convention-oriented behaviour as to which extent it is prevalent in society. Russell Hardin suggested that conventions lie behind much conformational behaviour of people within organizations, as they shirk the costs of breaking out of rules with a coordinative function and of finding new coordination, but he also speculates on the wider applicability of the coordination logic in society (Hardin 1998). His main example is state organizations, but in this case it is not easy to see how laws and roles come about coordinatively in the first place (Anderson 2000, 177). The approach to society from conventions, indeed, is reduced to cases where there is either a congruence of interests (as in right-hand driving) or where the compliance of norms is already in place and too costly to change. Even then divergence of interests may give people different views on their costs to 'recoordinate' when considering even long-term established norms (think, for instance, of the breaking up of a marriage). Overall, to reveal conventions as relevant for larger parts of society seems not to be successful.

²³ This proceeding is also oriented at the excellent overview given on recent literature in Anderson 2000. For a more detailed comment on the literature, as especially collected in Ben-Ner/Putterman (eds.) 1998, see her article, as well as Elster 1989.

'Assurance situations' are structurally broken down into the assurance game. The classical illustration of the assurance game is (referring back to Rousseau) the stag hunt. Two hunters could cooperatively hunt stags or singularly hunt hares. It is much better to come into the possession of a stag than a hare. But stags can only be hunted cooperatively, while hares can be hunted individually. So hunters may, during a hunt for stag that is not immediately successful, change their minds and grab a hare running by, thereby giving up the chances of attracting a stag. Strong-minded stag hunters are saddled, then, with neither stag nor hare, whereas the defectors are successful with a hare.

Again, there are rather specific conditions for cooperation involved in this assurance setting. On the one hand, in contrast to collective goods, the good aimed at cannot be achieved in part, but only absolutely. On the other hand there is a rational imbalance between two solutions for all contributors: they may either contribute and take the risk that their contribution (if not reciprocated) comes to nothing, or they may be successful (if reciprocated) and achieve a higher benefit. As there are two 'Nash-points' involved, the game encounters no stable solution, even if it does not push its participants towards the disastrous end of the prisoners' dilemma.²⁴

In theory '*prisoners' dilemma situations*' are thought to lead to cooperation only on condition of repeated interaction (with an infinite end), due to fear of retaliation and/or hope of reciprocation. Multiple improvements of the pure model arise from it. In the present context we should be interested to know how an ideally repeated PD-game maps onto real society. If repeated PD-games may be effective among a small group of participants (ideally two), it is difficult to see how they solve a typical public good-problem with a large size of potential contributors/defectors. To be more specific on this depends on the norm selected, as for example 'tit-for-tat'. This norm will work reasonably well among people in direct contact and controlling each other's behaviour. But in the bigger picture it is not clear how to identify defectors or from whom to receive punishment. In theory one would have to expect, then, that contributory behaviour in public goods deteriorates over time towards zero, something which surprisingly does *not* happen in real society to the extent expected (see Fischbacher/Gächter/Fehr 2001).

Recent evidence that actual behaviour is much more prosocial than predicted by public-choice theory challenges the basics of rational games. Why do people behave so differently as predicted from these games? Not many taxi passengers, for example, regularly attempt to not pay their bill, nor do many taxi drivers not try to follow and track down a cheating passenger, something they would be rationally advised not to do. Both passenger and driver seem not to behave as rationally as theory suggests and predicts, and the same seems true for many typical one-shot exchanges in normal life. And nor is this anti-rational behaviour based on anecdote; it has been systematically documented by many experiments with the one-shot ultimatum game, something reported in more detail in 4.3.

²⁴ See Camerer/Knez 1997 for further application that makes use of both conventions and the assurance game. More elementary is Skyrms 2001.

4.2 Cooperation on Communal Interests

The second class of potentially cooperative motivation may be formulated in terms of ‘communal interests’, but this term and the accompanied phenomena need to be explained further. One could characterize this class of motives as arising, in general, from an understanding of acting under a ‘we’. In analysing the whereabouts and the underlying logic of this ‘we’, many psychological details have to be taken into consideration: emotions and needs for social contact and stability, supraindividual values and aims, beliefs and intentions shared with others, identification and empathy, etc.

‘We’ attitudes, including ‘we’-intentions and ‘we’-beliefs, are to be distinguished from what was called ‘joint-intentions’ in the last section.²⁵ Intentions become ‘joint’ once there is a certain amount of interrelationship between individual intentions. For this the idea of a ‘we’ is not necessary, even if it is often helpful and sometimes the very source of joint intentions. Musicians in an orchestra may act in part from the basis of their being members of the orchestra; but most of their coordinated action can be analysed, and will actually be motivated, from the basis of their joint intentions, their shared common goal (playing) and their meshing work-plans. It is no wonder, therefore, that ‘we’-attitudes introduce in a stricter way the problem of how they are related to individual attitudes.

There seem to be, to begin with, two kinds of relation between individual agents involved in groups or collectives—like partners in marriage or business—and the collective agents themselves—like marriages and firms: the *reductively* individualist and the *unrestricted* collectivist kind of relation. In a reductivist understanding of ‘we’-relationships all we-attitudes will be constituted by an interpreted sum of individual attitudes, and by nothing else. For an unrestricted collectivist, on the other hand, individual attitudes in the sphere of collective matters become displaced by irreducible ‘we’-attitudes. Accordingly one can distinguish between reductively ‘individualist’ and unrestricted ‘collectivist’ ideas of ‘we’-attitudes. But it is important to take note of a third alternative not running into the deadlocks of both these extremes: the *compatibilist* reading.

The compatibilist reading accepts the irreducible and self-standing quality of social groups or collective agents. It accepts that it is something ‘in addition’ when two people speak from the point of view of their ‘we’ as opposed to their

²⁵ Unfortunately, terminology is confusing in the philosophical literature. Tuomela in particular is fond of the ‘we’ terminology, but does not distinguish clearly from what is here called joint intentions. See Tuomela 2007, ch. 1; 2011. Tuomela, in my impression, confounds (individual) joint intentions and group-consciousness, which he terms thinking in the ‘we-mode’. The confusion arises when he says “A1 and A2 intended jointly, as a group [...]” (2011, 73) Does he mean that joint intending *makes* a group or that only joint intending *as a group* is joint intending in an optimal sense, according to some normative criterion? He seems to incline towards the first (see also 72), but that is surely a definitional fiat which trivializes ‘group thinking’ and is not coherent with other remarks of his (‘ethos’, ‘group-identity’, etc.). To make visible the difference between joint actions and group actions: Tuomela is wrong when he says that “it is neither rational nor ‘group-socially’ beneficial for one of them to defect as long as the others do not defect” (2011, 74). There is a difference between contributing to a common goal and to a group’s goal, as proved by those individuals who try to save the firm’s name or the family’s reputation by not contributing, as the others do, to a bad goal.

corresponding ‘I’s, letting both either stand side by side or getting involved in conflict between them.²⁶ Overall the compatibilist variant definitely seems the more plausible, because it avoids the impasses of the alternatives and seems in good agreement with examples. Unfortunately, there is a catch to it. The alternatives include a clear form or rationality, be it individualist or collectivist, but the compatibility variant is unresolved in how individual and we-attitudes are to be correlated. There seems a plurality of relations possible: those of restriction and support, in different degrees, but perhaps even one of indifference. So far there has been, to my knowledge, no attempt to classify these possibilities.

For empirical research the compatibility variant in principle opens up a *two-level view* on motivation. In the case of a workers’ cooperative—as distinct to a family or religious community into which one is born—the material motive to earn one’s living may be a first and central motive to enter the cooperative, but in addition there may be further motives, springing from an interest in experiment, sympathy to just causes, solidarity with the workers’ tradition, a sense of freedom, etc., to ‘identify’ with the aims of the cooperative. In principle, the same may hold true for entry into a capitalist firm, with different content of the emotions and values involved. To talk of ‘justice’ or ‘democracy’ within the workers’ cooperative presupposes, of course, something like the compatibility idea of how workers and the cooperation are related to each other. The individual members in the cooperative have to remain, in part, self-standing, in order to qualify for the collective under these terms. Alternatively, there would only be global evaluations of the collective itself, referring to its strategy, business performance, success, etc.

Having examples like this in view, we may even suggest a four-sector model of motivation, encompassing both an individual and a collective part of our more desire/emotion-based motives and our belief/reflection-based ones. To simplify, it is possible to distinguish between individual and collective motives of a rational and reasonable kind. As plausible and perhaps also unavoidable as the dualism of desires and beliefs may seem, it involves the behaviour-focussed methodology of economic theories in extreme problems. There are indirect ways to get a grasp on intentions and beliefs, however, even if starting from a strict behavioural basis. I will touch upon some in 4.4.

To make use of the rational/reasonable distinction means that one adds the *dimension of beliefs* to rational desires, and this may also have the consequence that the I/we distinction is less important than is often believed, or at least less important ideally. To illustrate with the example of the workers’ cooperative: which motives could result from identification with the collective that are *not themselves* oriented at, or filtered through, evaluations? Does one really identify with a collective, or should one do it beyond individual desires and values? Values must be individually held, but are not individually restricted, as for example moral values in particular. The idea that the phenomenon of ‘plural subjects’ opens up a fully different dimension of motivation to the individual

²⁶ ‘We as a firm (marriage, university, committee etc.) have the view (decide to do, take the stand), that p, but I think we should rather take the view (should decide, take the stand), that q.’

(see, for example, Anderson 2000, 192, referring to Gilbert) thus seems a bit overstated. It is the distinction between the rational and the reasonable that does most of the work here.²⁷

Overall, the two-level view may take some steam out of overly enthusiastic views of ‘communal’ or even ‘communitarian’ potentials of motivation. To analyse such potentials with the help of a rational/reasonable-distinction may be the better normative approach. And I am somewhat convinced that Rawls is right in seeing beliefs of fairness and efficiency as ‘working in tandem’, and by this opening the opportunity of ‘fair cooperation’, but the mutual interdependence in this combination still awaits a precise analysis.

4.3 Altruistic Cooperation

This conjecture can be deepened by taking note of a good deal more of recent research that focuses on what seem to be ‘social’ attitudes beyond rational self-interest, involving partly moral ones. This research is more or less negligent of the I/we-distinction. In part it brings us back to the Knight-Buchanan-Gauthier dispute on the dependency of market motives on social/moral motives sketched earlier, but it is more recently based on provokingly new empirical findings.²⁸ To understand the messages of the experimental work involved it is again not irrelevant to be aware of the social/moral distinction, and by this again we are involved in the two-level view of motivation, including the rational/reasonable distinction.

‘Social’ motivation could be, analogously to ‘social action’, understood—even if now not on the level of belief but on the level of motivation—as motivation taking into account the interest of others. Social action can be partly or fully self-interested, as in action for social reward, but there are two further sources of motives which we normally think of when talking about ‘moral motives’: *love and altruism* (in the sense of identification with another’s welfare), and the motive of *duty*.²⁹ Even if both altruism and duty have the same object, another’s welfare, they spring from different departments of our psychology, from emotion and belief. Among philosophers there is the tendency already

²⁷ See, however, experiments demonstrating the relevance of group identification in groups based on pure chance. There seems to be an innate human tendency to associate with a group, more or less unfiltered by judgements. See: Kurzban 2003. The problems involved in such tendencies are obvious.

²⁸ Frey 1998 observed the ‘crowding out’ effect monetary compensation has on public waste-facility debates, reducing citizens’ readiness to decide matters in terms of fairness and equal rights. This could be read as a case of market relations being strongly in need of a consensus base that is differently motivated. (But think again about the underlying I/we effect: is being a Swiss citizen important for fairness or is not fairness involved in the conflict between Swiss citizens? I would incline towards the latter alternative.)—Frank 1998 analyses the conflict and mutual influence of rational behaviour and social norms, as, for example, norms against conspicuous consumption. He shows social norms, especially in small communities, to be instrumental for success in the economic domain.—Fehr and many others provided proof that workers behave in employment relationships in ways beyond self-interest, responding freely and fairly to higher pay within non-enforceable labour-contracts. See Fehr/Kirchler/Weichbold/Gächter 1998; Gächter/Fehr 2002 and the excellent overview in Fehr/Goette/Zehnder 2008.

²⁹ See the helpful discussion of these three motivational sources in Mansbridge 1998.

mentioned, convincing in the end, to restrict ‘moral’ motives to those of duty. Awareness of one’s duty (or moral rightness) may correct misguided emotional involvement out of love or altruism, and therefore is preferable to pure emotion (not to speak of the often contingently personal limitation of unreflected emotions).

The relevance of these distinctions again puts pressure on the methodologically restricted focus on observable decision-behaviour in recent empirical research. Without taking note of intentions the difference between altruism and duty cannot be detected (see Peacock/Schefzcyk/Schaber 2005). The experimentalists speak of ‘altruism’ only if there are costs to the agent and benefits to others, whatever the motives and beliefs involved. But at times, in a seemingly not fully coherent manner, they also speak not only of altruism but also of ‘fairness’-behaviour (Fehr/Gächter 2000), a distinction to be made only in the light of intentions and motives. A deeper interpretation of motives seems unavoidable.

To be fair to this research itself, the methodological problem is not as intractable as shown so far. Even if the basic imputation of fairness motives in the context of purely behavioural data may be shaky,³⁰ there is some amount of indirect proof of the presence of fairness considerations implied in behaviour. This pertains especially to behaviour in experimental games and behaviour in the labour market and in firms, both methods of research being in an experimentally supportive relationship to each other. Fairness concerns can be detected in work relations analogously to the ultimatum game, relations between agents in the game and the labour contract being highly similar. Offers by proposers in the games are responded to in accordance to their fairness. Analogously, in contractually incomplete employment-relations, workers respond by choosing different effort levels to different wage offers. Similar to the ultimatum game, there are a number of completely selfish workers not responding to ‘fair’ offers. But also similar to the ultimatum game, a small number of fair-minded workers invoke a potentially punishing effect to these selfish workers. Even if selfish, people are often reputation-seekers and therefore open to conformity pressures.

These findings, within the important sphere of work and production, give further comment to several points already touched upon. First, they offer an illustration to which extent social (in the present case economic) relations determine in part the kind of motivation appropriate for a relation. Labour markets do not fit easily into the RCM, because work efforts cannot be as narrowly contracted as the exchange of consumer goods. Whereas in the exchange of the latter there is small room for nonrational behaviour, in typical work-relationships—as in many social relationships outside the classical market—there is scope for

³⁰ “A plausible interpretation for the rejection of low offers [in the ultimatum game]—which is typically also supported by responders’ verbal accounts of their behaviour—is that subjects perceive them as unfair.” (Fehr/Goette/Zehnder 2008, 5)—Indirect proof of intentions also comes from game-settings that restrict the options of the proposer. Responders decide differently if confronted with offers that are restricted or not restricted for the proposer. Knowledge of the proposers’ range of options to offer influences judgements of the degree of fairness of the offers. See *ibid.* 8.

the development of nonrational motives. And perhaps more: these motives are needed for work-relationships to prosper.

Secondly, actual behaviour indeed proves that a certain amount of people are behaving according to a dual-level model of motivation, accepting costs in rational terms on behalf of fairness-motivation, or on behalf of their reasonable side. Thirdly, a longer history of ongoing relations and the reciprocity effected in them contributes positively to the fairness part of motivations. It also forces the selfishly motivated agents into increasing efforts of fairness-mimicking behaviour.

Given our thoughts on we-intentions in the last section, these findings give no proof to them so far. There is some proof of reference-relativity for fairness responses: workers react differently if their wage is judged from outside or inside the firm, depending on whether they are beginners or long-term employees (Fehr/Goette/Zehnder 2008, 26–31). But it is not obvious here that anything more than comparison frames necessary for judgements of fairness are in play. Of course, this research did not look explicitly for potential social group effects behind its fairness motives; but on the other hand none seem to impress themselves on the experimentalists. Collectives as socially defined spheres of action are, however, important in another sense: only within a circumscribed collective—understood here as a circumscribed sphere of action—is punishing possible, and without punishing the fair-minded agents would not have the collectively positive effect they have.

4.4 Cooperation through Norms

Lastly, I will take a short look at the potential role of social norms for cooperation, both on the basis of rational and moral motivation. It is, of course, impossible to make any substantial remarks on this complex topic, given the restrictions of space. But perhaps we can see, even briefly, a complementary tendency of argument to the ones outlined so far.

Why are social norms such a complex topic to begin with? For analysis, the basic reason is again the one showing up in every attempt to understand social action: the unavoidable interdependence of concept/definition and explanation. (And the same is good for all acts of understanding, which leave the everyday certainty of concepts behind.) The explanatory role one is inclined to think that norms have determines meaning, and vice versa. To disentangle this interdependency makes any debate about social norms a laborious affair. We can follow further, however, the tracks already laid so far and consider first how the rational approach pans out, and secondly what the alternatives are.

The unavoidable interdependency of concept and explanation typically shows up in reporting, even if roughly, the two alternative understandings of norms: norms as *sanctions* invoked by others—corresponding to rational explanation—and norms as *internalization*—corresponding to sociopsychological explanation. These are, again roughly, the economic and sociological traditions involving, first, how norms are understood, and secondly how norms work; it would be fruitless to fully separate meaning and explanatory intention as given in these packages. Explanation is done by *reduction* in the first case, and by *descrip-*

tion in the second. One explains externally, the other internally or by a whole (behaviour)/part (dispositions) relation. The two perspectives on norms differ, because theorists want norms to do different work, given different descriptions of agents, their basic characteristics and problems the agents involve themselves in.

If rational agents are not able to cooperate robustly in prisoners-dilemma-like situations, can norms help them out?³¹ On the basis of the rational approach, ‘norm’ would stand here for expectations among agents to be sanctioned by others if not cooperating. In the simple case of two agents, this would have to be third-party people or the social environment. If rational agents have a justified belief that possible sanctions by others are more costly than the benefit to be reaped through non-cooperation, they will cooperate. There is a series of reasons, however, that the readiness to sanction cannot stand on rational motivation alone. The basic objection is this: if all others are also rational agents, they are part of an extended prisoners’ dilemma, and the question is open as to why they should take the burden of sanctioning others without reaping benefits from it. Norms are a public good and cannot be explained by a cooperative logic, which is unable to explain public goods at the more elementary level.³²

The upshot of this failure is that the understanding both of norms and the underlying motivation has to be changed. In order to escape the impasse, either self-interested motivation has to be given up, or the primary role in norms of sanctioning, or both. It makes no sense to simply enlarge self-interest by adding norm-relative content that then is excluded from the explanation.³³ The alternatives come, in principle, from a different, non-self-interested motivation in combination with a social interdependency, either on the basis of emotions or beliefs, or both. To see these alternatives more clearly is complicated by the insight that self-interest will still have to be part of a more complex conception of norms, as the postulation of non-self-interested motives throughout would be confronted with the question of why norms are needed in the first place. That much, however, is clear: the rational approach is unsuccessful in its attempt to close its loopholes by falling back on norms.

What is not clear, unfortunately, is how a more descriptive or internal (and not reductive) explanation of norms is meant to work. The interesting thing to note is that sanctioning from non-self-interested motives is documented as being effective, and effective to some extent in a quite general way, not bound to specific group-involvements.³⁴ Fehr and Fischbacher (2004; 2004a) showed in exper-

³¹ For the time being let ‘norms’ be understood as ‘social norms’; further remarks on content will follow shortly.

³² For more detailed argument critical towards the rational approach on norms see also Elster 1989, ch. 3; 1989a; 1990, 872–5; Anderson 2000, 181–4. There are many variants of how the rational approach could be supplemented or situated that I have to neglect here. Also I neglect altogether evolutionary theories of social norms. For critical comment on the latter see Anderson 2000, 184–91.

³³ “Whenever a norm is invoked in explaining an action, one can turn the account into a rational-choice explanation by saying that the agent is concerned with upholding his self-image or reputation as a norm-follower. While true enough, such statements are uninformative.” (Elster 1990, 872)

³⁴ An observation that weakens the relevance of typical ‘we-motives’ springing from involve-

iments that there is roughly up to 60% readiness by third-party players to punish violations of what they describe as violations of fairness in dictator- and public goods games. The motivation seems clearly non-egoistic, as third-party punishers do not have the possible motive of revenge that is perhaps active in second-party players, and also as they forsake money for punishing. Third party-players seem to reveal “the truly normative standards of behaviour” (Fehr/Fischbacher 2004, 65). The explanatory deficit of these experiments is, unfortunately, that they do not let us see why people behave like this. Fehr/Fischbacher and many other experimentalists seem to assume endogenous motivations at work: in this case a sense of fairness triggering readiness for costly sanctioning.

One can see how this falls short of the phenomena if one contrasts readiness for sanctioning with readiness for ‘right’ behaviour to begin with. If nothing else but a sense of fairness would be at work, why are so many players in need of being sanctioned? The propensity of sanctioning seems conspicuously higher than the propensity to comply, which makes a simply endogenous explanation incoherent. Observations of asymmetry between compliance and sanctioning among agents of otherwise similar stature not only disqualifies classical rational explanations of norms, but also explanations which substitute rational motives by emotional ones of various sorts: feelings of fairness, sympathy, indignation, disgust, etc. What these explanations neglect is the constitutive function for norms of different *roles within social relationships*, as especially seen in the roles of agent, co-agent, bystander, wider public, etc.³⁵ Sanctioning seems to be necessary for resilient cooperative relations, but to explain this further we have to take into account the priority of *other-related* dispositions.

Helpful in the context of these observations may be the distinction between *other-dependent* emotions, like self-esteem, pride or envy, and *other-directed* emotions, like punishment or disapproval, and the acceptance of the constitutive role these qualities in humans have. Explanation of norms, which tries to start ‘below’ the other-relatedness of these emotions, remains in danger of disobeying the function social constellations *per se* have for human behaviour. This creativity of the social constellation can hardly be explained by deeper-laying, not other-related tendencies. Social norms are only to be understood *out of* emotionally structured relationships, and should not be seen as externally constitutive for relationships. Whether this is a definitive breach with methodological individualism is to be considered; it certainly asks for a reinterpretation, as agents now have to be approached basically in their *being related to* others, even if it is single agents so related.

ment in partnerships, groups and collectives. Certainly such involvements may strengthen normative dispositions, but they cannot launch them on their own.

³⁵ For highlighting this point see Anderson 2000, 184, 193–5. Sociologists will, from their awareness of the psychological mechanisms of ‘interaction’, approve of this diagnosis. But up to now they have not made ample use of the truly fine-grained analyses of social interdependencies offered by analytic philosophers. These analyses by far surpass what the ‘symbolic interactionists’ thought to be at work in social dyads.

5. Making Cooperation Work

Two things seem to be most striking in this overview. On the one hand the extent to which the concept of ‘collective action’ and ‘cooperation’ are invoked differently by the two schools of social theory. On the other hand, and perhaps corresponding in a way, the failure of rational theories in the narrowest sense to model something at least nearly similar to actual behaviour. But contrasting everyday understanding of cooperation (‘collective action’ is a less regular word) as made explicit by philosophers opens up many ways for the experimentalist of how to interpret actual ‘cooperative’ behaviour and perhaps to extend experiments accordingly. With recent insight into the fine-grained intentional interdependencies of ‘cooperation’ and the meanwhile quite explicit evidence of fairness motivation, given suitable (especially reciprocal) conditions, it should be possible to integrate these two developments in a sense that lifts the experimentalists’ theoretical framework further from its behavioural bases.

Among the most urgent open questions, to my mind, are two in particular: the role of moral motives (as distinguished from altruist ones), and the role of being in relation with others *per se*. As far as the role of moral motives is concerned, the theoretical gap between moral philosophers, authorized for the morally right, and social theorists is depressingly wide. It is difficult to understand, from the side of the latter, why the philosophers are simply not interested in the *actual extent* of moral motivation and behaviour in normal life! Or, to put it a little more mildly, why philosophers are so much more interested in the intellectual side of morality than in its motivational side. Unfortunately, given even the recent research covered here, it remains unclear as to which extent it is ‘moral’ motives at work, being mindful also of the different senses of ‘moral’.

Elster once illuminatingly illustrated the different behaviour of Kantians, utilitarians and altruists in the development of a collective project (Elster 1985, 148–152; 1989, 34–48, 202–214). What shows up from this is how ambiguous, and at times harmful, the unconditional behaviour of Kantians could be, if understood as a general motive and not a subsequent reason for a ‘moral’ situation independently defined. Many collective projects would not get started without agents who are unconditionally motivated, as Kantians are. What Elster refers to as “everyday Kantianism” is itself a mixture of different habits and thoughts, and not solely expressive of the Categorical Imperative. It would be interesting to see how far this behaviour converges with that studied by the recent experimentalists. In Elster’s illustrations, too, reciprocal cooperation seems to play a role, but it remains unclear as to which one it is. Are his Kantians heading towards a cooperative goal or are they non-cooperative, something not inconsistent with being Kantian?³⁶

Also seemingly obvious is the impression that self-interested cooperation and fairness, the rational and the reasonable, can go hand in hand, based on some actual amount of reciprocal cooperation. Such a dependency is not harmonious with unconditional motivation, and therefore is morally more restrictive. On

³⁶ According to the distinctions made in figure 1.

the other hand, the idea of cooperation may be extended to rather abstract and indirect forms of (potential) cooperative relations. If our social dispositions are, as it seems, strongly *other-related*, other-dependent and other-oriented morality, law and politics would not be possible without a certain amount of effective and mutually satisfying cooperation. Norms in these social spheres seem to be the product of cooperation, and not the other way around.

The most shockingly bad things for humans result from other humans. Humans are exceptionally creative in harming other humans. But being positively related to others is also a deep human characteristic, and if we apply again the test of universalization, it must normatively be the primary one—otherwise humans would not exist anymore. Given this inter-individual creativity, cooperation seems to be the most promising phenomenon to work on for answering the problem of order, as complex as, finally, such an answer may be.

Bibliography

- Anderson, E. (2000), Beyond Homo Economicus: New Developments in Theories of Social Norms, in: *Philosophy & Public Affairs* 29(2), 170–199
- Ben-Ner, A./L.Putterman (1998) (eds.), *Economics, Values, and Organization*, Cambridge
- Bratman, M. (1992), Shared Cooperative Activity, in: *Philosophical Review* 101(2), 327–341
- (1993), Shared Intention, in: *Ethics* 104, 97–113
- Buchanan, J. (1975), *The Limits of Liberty*, Chicago
- (1988), The Gauthier Enterprise, in: *Social Philosophy & Policy* 5(2), 75–94
- Camerer, C./M. Knez (1997), Coordination in Organizations: A Game-theoretic Perspective, in: Shapira, Z. (ed.), *Organizational Decision Making*, Cambridge, 158–188
- Coleman, J. (1987), Competition and Cooperation, in: *Ethics* 98, 76–90
- Davidson, D. (2001), *Subjective, Intersubjective, Objective*, Oxford
- Elster, J. (1982), Marxism, Functionalism, and Game Theory, in: *Theory and Society* 11, 453–482
- (1985), *Making Sense of Marx*, Cambridge
- (1985a), Rationality, Morality and Collective Action, in: *Ethics* 96, 136–155
- (1989), *The Cement of Society*, Cambridge
- (1989a), Social Norms and Economic Theory, in: *Journal of Economic Perspectives* 3(4), 99–117
- (1990), Norms of Revenge, in: *Ethics* 100, 862–885
- Fehr, E./E. Kirchler/A. Weichbold/S. Gächter (1998), When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets, in: *Journal of Labor Economics* 16(2), 324–351
- /S. Gächter (2000), Fairness and Retaliation: The Economics of Reciprocity, in: *Journal of Economic Perspectives* 14(3), 159–181
- /U. Fischbacher/S. Gächter (2002), Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms, in: *Human Nature* 13, 1–25
- /U. Fischbacher (2004), Third-Party Punishment and Social Norms, in: *Evolution and Human Behaviour* 25, 63–87
- / — (2004a), Social Norms and Human Cooperation, in: *Trends in Cognitive Science* 8(4), 185–190

- /L. Goette/Ch. Zehnder (2008), *A Behavioral Account of the Labour Market. The Role of Fairness Concerns*, IZA Discussion Paper Series 3901, Bonn
- Fischbacher, U./S. Gächter/E. Fehr (2001), Are People Conditionally Cooperative? Evidence from a Public Goods Experiment, in: *Economic Letters* 71, 397–404
- Frank, R. (1998), Social Norms as Positional Arms Control Agreements, in: Ben-Ner/Putterman (eds.), ch. 10
- Frey, B. (1998), Institutions and Morale: The Crowding-out Effect, in: Ben-Ner/Putterman (eds.), ch. 17
- /F. Oberholzer-Gee/R. Eichenberger (1996), The Old Lady Visits Your Backyard: A Tale of Morals and Markets, in: *Journal of Political Economy* 104(6), 1237–1313
- Gauthier, D. (1986), *Morals by Agreement*, Oxford
- Gächter, S./E. Fehr (2002), Fairness in the Labour Market. A Survey of Experimental Results, in: *Surveys in Experimental Economics*, Part II, 95–132
- Gilbert, M. (1989), *On Social Facts*, London (2nd ed. 1992, Princeton)
- (1996), Fusion, in: *Living Together*, Lanham–New York–London
- (2008), Two Approaches to Shared Intention: An Essay in the Philosophy of Social Phenomena, in: *Analyse & Kritik* 30, 483–514
- Hann, C. (2006), The Gift and Reciprocity: Perspectives from Economic Anthropology, in: Kolm, S. C./J. M. Ythier (eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Vol. 1, Brussels
- Hardin, R. (1998), Institutional Commitment: Values or Incentives?, in: Ben-Ner/Putterman (eds.), ch. 16
- Knight, F. (1951), The Role of Principles in Economics and Politics, in: *The American Economic Review* 41(1), 1–29
- Koslowski, P. (1982), *Die Ethik des Kapitalismus*, Tübingen
- Kurzban, R. (2003), Minimal Group Experiments, in: *The Encyclopedia of Cognitive Science*, London
- MacDonald, G./P. Pettit (1981), *Semantics and Social Science*, London
- Mansbridge, J. (1998), Starting with Nothing: On the Impossibility of Grounding Norms Solely in Self-Interest, in: Ben-Ner/Putterman (eds.), ch. 5
- Parsons, T. (1937), *Structures of Social Theory*, New York
- Peacock, M./M. Schefczyk/P. Schaber (2005), Altruism and the Indispensability of Motives, in: *Analyse & Kritik* 27, 188–197
- Rawls, J. (1993), *Political Liberalism*, Cambridge/MA
- Schaber, P. (2011), Was moralische von altruistischen Motiven unterscheidet, in: Schmidt, T./T. Tarkian (eds.), *Naturalismus in der Ethik*, Paderborn, in print
- Skyrms, B. (2001), The Stag Hunt, in: *Proceedings and Addresses of the American Philosophical Association* 75(2), 31–41
- Tuomela, R. (2007), *The Philosophy of Sociality. The Shared Point of View*, Oxford
- (2011), Cooperation as Joint Action, *this issue*