

**Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2013/21**



Fairness and Reciprocity
in Contract Governance

Stefan Magen



MAX PLANCK SOCIETY



Fairness and Reciprocity in Contract Governance

Stefan Magen

October 2013

Fairness and Reciprocity in Contract Governance

Stefan Magen¹

Abstract

Contract Governance is, amongst others, concerned with solving problems of cooperation and opportunistic behaviour. However, with *homo reciprocans* as a party to the contract, contractual relationships are necessarily accompanied by expectations of fairness and reciprocity. In this regard, the last two decades have witnessed tremendous progress towards a behavioural account of fairness and its pivotal role in sustaining cooperation. In this article, I sketch some important features of the general working mechanisms of fairness based on a selective review of the literature.

It is crucial to distinguish mental, behavioural and social aspects of fairness and to see how these work together. Experimental game theory has focused mainly on the behavioural level and developed two types of models of preferences for fairness: inequity aversion, concerned with the distribution of outcomes, and reciprocity, concerned with the fairness of intentions. Both assume preferences for fairness to interact with selfish preferences that remain a primary motivational force. On the mental level, however, fairness judgments also allow for several proportional criteria of distribution, as Equity Theory in social psychology has shown. The (psychological) necessity to select among different criteria of fairness and their operationalization renders fairness judgments indeterminate and makes them receptive for social and contextual influences. Because of this, *homo reciprocans* is a rather impressionable being and his behaviour heavily depends on the specifics of the situation and the social context in which he is acting.

Compared to *homo oeconomicus*, cooperation is both easier and more difficult with *homo reciprocans*. It is easier because reciprocity repays cooperation with cooperation. It is more difficult because opportunistic behaviour by selfish types, usually mixed into real-life popula-

1 Ruhr -University Bochum, Chair in Public Law, Legal Philosophy, and Law and Economics.

tions, triggers negative reciprocity on the side of reciprocal actors causing a downward spiral that eventually extinguishes cooperation. Furthermore, even fairness judgments are unconsciously biased toward self-interest if and insofar as factual knowledge or normative standards are ambiguous. Sanctions therefore remain crucial to sustain cooperation, but they assume quite a different objective compared to *homo oeconomicus*. With *homo reciprocans*, sanctions have to provide social or moral incentives in order to sustain positive reciprocity, and they do so indirectly by preventing opportunism and thus keeping social relations fair or balanced. Negative reciprocity provides sufficient motivation to leave punishment to the parties, but excess punishment and anti-cooperative punishment, combined with self-serving perceptions, call for institutional intervention in order to prevent a vicious circle. Although people in general show an aversion against punishment institutions, they are accepted as legitimate if they are chosen autonomously. Furthermore, because of the indeterminacy of fairness, there is a need to settle on shared fairness standards, if cooperation is to be based on positive reciprocity and protected against negative reciprocity. This creates a higher-order coordination problem, which institutions can help to solve, because fairness judgments intrinsically rely on reference points that institutions can provide. In Contract Governance, both statutes and contracts can serve this purpose through their expressive effect. Also, the mere fact of making a promise and the 'procedure' of doing so with reference to the law induces a commitment to the contract. In short, Contract Governance has to tame and institutionalise sanctions and it has to coordinate fairness standards via the expressive dimension of norms and contracts.

Fairness and Reciprocity in Contract Governance

Contract Governance, as it seems, examines a set of rather different regulatory constellations.² In the analysis of these, fairness or reciprocity would be of relevance in quite different ways that cannot be pursued here. Instead, I take governance for the purpose of this article in a narrower and at the same time more general meaning that cuts across the different forms of Contract Governance. In the following, I understand governance to be occupied (among other things) with problems of cooperation or opportunistic behaviour, in which mutually beneficial outcomes are not realized because actors have an incentive to deviate from cooperation or, put differently, in which the efficient outcome is not an equilibrium.³ The canonical model for this type of social problems is, of course, the Prisoner's Dilemma, but in fact cooperation problems take a variety of forms that pose different problems, for the analysis of which a multitude of more sophisticated and more adequate models are being applied. The following article is interested in the role of fairness or reciprocity in solving problems of this sort and it focuses more on general mechanisms than on specific results (whose external validity has to be considered with caution and care).

I What is fairness or reciprocity?

Before we look at what fairness or reciprocity entails, it is important to recognise that this question can be asked at different levels: at the mental, at the behavioural and at the social level. At the mental level, fairness can, for example, take the form of moral sentiments or of moral judgments; on the behavioural level, for example, the form of voluntary cooperation or of a rejection of inequitable offers; and on the social level, for example, the form of shared normative expectations (an aggregate of mental phenomena), of a pattern of sanctioning behaviour (an aggregate of behaviour), or of a combination of both aggregate behaviour and expectations. Which level one should focus on depends, of course, on the question to be pursued and should be delegated to different disciplines accordingly. With regard to the analysis of governance problems, it seems helpful to adopt a more comprehensive approach and to conceive of fairness as a "social mechanism" to sustain cooperation in human interactions based on individual behavioural disposition. From this point of view, the different levels mentioned above refer to different aspects of the same phenomenon, which are somewhat independent from each other, but which also interact. Thus it becomes crucial to understand how the different – mental, behavioural and social – aspects of fairness work together in bringing

2 F. Möslin / K. Riesenhuber, 'Contract Governance – A Draft Research Agenda', *European Review of Contract Law* (2009), 248, 260 – 287.

3 This understanding should be covered by the use of the term by Elinor Ostrom and Oliver Williamson; see E. Ostrom, *Governing the commons* (Cambridge, UK: Cambridge Univ. Press, 1990); id., *Understanding institutional diversity* (Princeton: Princeton Univ. Press, 2005); O. E. Williamson, 'The New Institutional Economics', *Journal of Economic Literature* 38 (2000), 595, 601; id., 'Transaction Costs Economics', *American Economic Review* 100 (2010), 673, 679.

cooperation about, what their respective contributions are, how there are linked with each other and what difficulties arise in their interplay.

1 The identification problem: moral motive or selfish strategy?

This, however, is not easy because any account of fairness is faced with serious identification problems. A similar pattern of observable behaviour can result from different types of motivations (e.g., selfish motives, moral motives based on ethical reasoning or on moral emotions, or simply the motive to follow a rule that is regarded as appropriate) and individual motivation can be differently impacted by social influences (e.g., via hard sanctions, refusal to cooperate further, social disapproval, role models or information about the behaviour of others or about social expectations).⁴ Consider the case of a supplier in the automotive industry who has the chance to produce parts at a lower cost, but with lower quality and a slightly higher probability of deterioration. Although there is only a very small probability of detection, the supplier does not seize the opportunity to produce inferior quality and trade it for normal quality. If we focus on the observable behaviour, the story suggests an instant of fair behaviour, but in fact there might be quite different underlying motives. For example, business relations between the producer and the customer might be permanent and quite valuable or even vital for the supplier, and given the high stakes, even with a small probability of detection the expected loss might outweigh the gains to be expected from cheating. The point here is that it is often difficult to make valid inferences about the selfish or moral character of a motivation from behaviour alone. Even if we knew all relevant circumstances and could determine that behaving fairly was the best strategy to maximize income, the actor could still have chosen this behaviour out of moral concerns. The same can be the case the other way round. There are situations where people's behaviour seems to be selfish, whereas their actual motivation is that they feel unfairly treated. Imagine an employee who is shirking his obligations at work. This might be because the employee knows that the employer cannot observe the quantity or quality of his work and thus minimizes his effort, or because his sense of justice is hurt because colleagues are paid higher wages (which should not, as such, bother a rational utility maximizer who is concerned only with his own payoffs). If, however, fairness and selfishness are so easily confused, then why should Contract Governance bother at all and try to tear them apart? The reason is that institutions have to achieve considerably different tasks, depending on whether people are solely driven by selfish or also by moral motives. Thus, with fair-minded people, institutions have to be designed differently.

2 *Homo oeconomicus* vs. *homo reciprocans*

Since mental phenomena can never be directly observed (not even with brain-imaging techniques) and self-reports are notoriously unreliable, it is crucial to study the impact of fairness

4 For an overview of the psychological debate on egoism, see, e.g., C. D. Batson, *The altruism question: Toward a social-psychological answer* (Hillsdale, NJ: Erlbaum, 1991).

motives on behaviour in diagnostic situations, i.e., in situations where it can be precisely determined which actions are in the self-interest of an actor and which are not. The scientific tool that allowed for enormous progress in this regard was laboratory experiments that implement games as defined in game theory (e.g., the famous Ultimatum Game).⁵ These experiments track participants' investment behaviour in games where real money is at stake, comparing it to the game-theoretic prediction what a rational self-interested agent would do to maximize his outcome.

Two consistent behavioural patterns thus emerged from numerous experiments conducted in a multitude of variations,⁶ many of which were carried out by Ernst Fehr or are inspired by his seminal work.⁷ In what economists call "theories of fairness", these two behavioural tendencies have been formalized mathematically as two types of preferences that have been called inequity aversion and reciprocity. However, as one has to bear in mind, economic theories of fairness substantially differ from psychological theories on the same subject. Contrary to the latter, the former have no intention realistically to map the cognitive processes that eventually lead to fair behaviour. Rather, they try to strip the complexity of mental phenomena down to those factors that are necessary to capture the general behavioural patterns resulting from moral motives. Put differently, economic fairness theories do not aim to explain fairness. Rather, they model fairness in order to explain economic behaviour. Thus, the concept of *homo reciprocans*, which behavioural economists have put forward,⁸ is just as well an artificial model of human behaviour concocted for the purposes of economic theorizing as was the old *homo oeconomicus* of neoclassical economics, whom *homo reciprocans* is meant to replace.

How do both models differ? Whatever preferences *homo oeconomicus* is assumed to hold, it is usually required that they conform to the assumption of selfishness, which is the presumption that people are indifferent about the utility of others and care only about their own consumption. This rules out all types of other regarding preferences or "social preferences", e.g., altruism, spite, envy, competitiveness and fairness.⁹ Preferences for fairness, on the other

5 For a discussion of the experimental method and its methodological merits and problems, see, e.g., S. Gächter / C. Thöni, 'Micromotives, microstructure, and macrobehavior', *Journal of Mathematical Sociology* 35 (2011), 26; F. Gula, *The Methodology of Experimental Economics* (Cambridge, UK: Cambridge Univ. Press, 2005), 141 et seq.; S. Levitt / J. List, 'What do laboratory experiments measuring social preferences reveal about the real world?' *Journal of Economic Perspectives* 21 (2007), 153. For a discussion of the application of experimental findings for policy-making and in the law, see, e.g., H.-T. Normann / R. Ricciuti, 'Laboratory experiments for economic policy making' *Journal of Economic Surveys* 23 (2009), 407; K. Zeiler, 'Cautions on the Use of Economic Experiments in Law' *Journal of Institutional and Theoretical Economics* 166 (2010), 178; commented on by C. Engel, 'The multiple use of experimental evidence in legal scholarship', *ibid.*, 199; id., 'The difficult reception of rigorous descriptive social science in the law', in N. Stehr and B. Weiler (eds), *Who owns knowledge?* (New Brunswick, NJ: Transaction Publishers, 2007), 169.

6 For an overview, see, e.g., E. Fehr / H. Gintis, 'Human motivation and social cooperation: Experimental and analytical foundations' *Annual Review of Sociology* 33 (2007), 43.

7 Starting with E. Fehr / G. Kirchsteiger / A. Riedl, 'Does fairness prevent market clearing? An experimental investigation' *Quarterly Journal of Economics* 108 (1993), 437.

8 E. Fehr / S. Gächter, 'Reciprocity and economics: The economic implications of Homo Reciprocans' *European Economic Review* 42 (1998), 845, S. Bowles / H. Gintis, 'Homo reciprocans' *nature* 415 (2002), 125.

9 J. Sobel, 'Interdependent preferences and reciprocity' *Journal of Economic Literature* 43 (2005), 392.

hand, are defined by departing from the selfishness assumption in a specific way. Preferences for fairness are social preferences in the sense that they also consider what other people get (be it more or less). However, unlike altruistic preferences, preferences for fairness are not about other people's welfare as such, but about relative differences in gains or losses between the actor and other people. That is, fairness is not about A doing B something good, but about whether A gets more or less than B, and for what reason.

Inequity aversion – the first type of fairness-preferences – takes fairness as being concerned with outcomes, or more specifically with how outcomes are distributed. It assumes that payoff differences have, as such, a negative utility for those involved, and that this disutility is greater for disadvantageous rather than for advantageous inequalities.¹⁰ Take, for example, an inequality-averse player in the famous Ultimatum Game, who is faced with a proposal that concedes him 2 out of 10 € and leaves the proposer with 8 €. In his decision whether to accept the offer or not, this player would not only take the positive utility into account that he would derive from getting 2 € but also the negative utility he would experience from getting 6 € less than the proposer. Depending on the strength of his aversion against inequality, this negative utility might outweigh the positive utility from getting at least some payoff. Thus, the assumption that having something is always better than having nothing does not necessarily hold for inequality-averse players. Instead, to avoid the displeasure of being treated unfairly, such a player might reject a positive offer, as can be observed in experimental games all over the world.¹¹ But people are not only willing to forgo considerable amounts of money to avoid unequal distributions to their disadvantage. They also tend, albeit to a lesser degree, to share equally to avoid the negative utility from having an unfair advantage.

Theories of reciprocity, as the other type of fairness theories, do not refer to unequal outcomes directly, but assume that people primarily care about the intentions of other actors, i.e., whether they are treated kindly or not. Hence, in these theories the fairness of a distribution is not relevant as such, but indicates the fairness of an actor's intentions.¹² Like in theories of inequity, it is assumed that people suffer disutility from being treated unfairly or treating someone unfairly. But it is also assumed, that they gain positive utility from responding with kindness to kind people and with unkindness to unkind people. This provides a more direct explanation for why participants in experimental games are willing to reciprocate fair or kind or cooperative actions even if they have to pay a price for this or could have taken advantage of (like in the above example of the supplier of automotive parts). The tendency to repay at a

10 E. Fehr / K. Schmidt, 'A theory of fairness, competition and cooperation', *Quarterly Journal of Economics* 114 (1999), 817; G. Bolton / A. Ockenfels, 'A theory of equity, reciprocity, and cooperation' *American Economic Review* 90 (2000), 166.

11 Henrich et al., *Foundations of Human Sociality* (Oxford: Oxford Univ. Press, 2004); Henrich et al., 'Costly punishment across human societies', *Science* 312 (2006), 1767 ff.

12 M. Rabin, 'Incorporating Fairness into game theory and economics' *American Economic Review* 83 (1993), 1281; G. Charness / M. Rabin, 'Understanding social preferences with simple tests' *Quarterly Journal of Economics* 117 (2002), 817; M. Dufwenberg / G. Kirchsteiger, 'A theory of sequential games' *Games and Economic Behavior* 47 (2004), 268; A. Falk / U. Fischbacher, 'A theory of reciprocity' *Games and Economic Behavior* 54 (2006), 293. For empirical evidence, see A. Falk / E. Fehr / U. Fischbacher, 'On the nature of fair behavior' *Economic Inquiry* 41 (2003), 20.

cost is particularly pronounced when it comes to punishing someone by whom one has been treated unfairly.

Both types of fairness theories do not pretend that people solely care about fairness. Rather, they take the utility from the actor's own payoff as a primary motivational force, but expect this motivation to be mediated by fairness to some degree. Thus, the interesting question arises of how fairness does so. What outcomes result from the interplay of fair and selfish motives depends, of course, on the specific situation. Here, the behavioural game-theoretic approach is particularly illuminative because it helps to elucidate how or in what ways different motives work together in particular strategic situations to produce certain social dynamics that lead to, say, the unravelling of cooperation in a contractual relationship or in the provision of public goods, of which we shall later speak more. Before addressing this question, we should consider the conceptual limitations of purely economic accounts of fairness.

3 Equity Theory

As said before, economic theories model fairness for the purposes of economic theorizing. They aim at aggregate outcomes, but do not intend to provide a realistic account of the underlying cognitive or psychological processes. This seems a reasonable approach for many economic questions, but in order to understand the significance of fairness motives for governance problems and the way fairness motives impact on them, one has to move beyond. Three related issues are of particular importance here. First, it does matter how distributional concerns (inequity aversion) and relational concerns (reciprocity) relate to each other, but this relationship remains unclear in economics. Second, people do not oppose unequal distributions in any case, but are willing to accept differences in outcomes if there are legitimate reasons to differentiate. It is a feature of fairness judgments, first described by Aristotle¹³ and well established empirically by social psychological Equity Theory, that just distributions also include distributions proportional to some criterion of 'worthiness' or 'desert', in particular proportional to merit (esp. contributions) or need.¹⁴ On the surface, this equitable nature of fairness judgments has been recognized in behavioural economics and indeed demonstrated in many experiments. In fact, outcome-oriented fairness theories have been called theories of *inequity*. However, proportional distribution criteria have hardly found their way into these preference models, presumably to keep the models manageable.¹⁵ Third, acknowledging that fairness can also mean, and in fact very often does mean, a proportional distribution according to some criterion of worthiness brings to the fore another feature of fairness judgements that has plagued philosophical ethics for centuries, i.e., the indeterminacy of justice. When fair-

13 In Book Five of the *Nicomachean Ethics*.

14 D. Messick / K. S. Cook, *Equity Theory* (New York: Praeger, 1983); E. Walster / G. W. Walster, 'Equity and social justice' *Journal of Social Issues* 31 (1975), 21; T. R. Tyler et al., *Social justice in a diverse society* (Boulder, Col.: Westview, 1997), 45.

15 W. Güth, 'On inequity aversion' *Papers on Strategic Interaction* 524 (2005) (Max-Planck-Institute for Economics: Jena). But see R. Selten, 'The equity principle in economic behavior', in H. W. Gottinger and W. Leinfellner (eds), *Decision Theory and Social Ethics* (Dordrecht: Reidel, 1978), 289.

ness can assume several forms, the question arises which criterion of justice is to be preferred in general or in particular situations (equality, merit, need, status or rights?) and how it is to be operationalized and applied. In fact, much of the ethical or political debate is occupied with these questions. Each individual who articulates a preference based on concerns for fairness is faced with similar structural problems. But in everyday decisions, people rarely solve this problem by deliberate principled reasoning the way philosophers do. They rarely even are aware of it, but rather form their social preferences intuitively. It is no surprise that fairness preferences often do not behave as rationally as is required by economic theory. Instead, they are strongly influenced by personality traits, peculiarities and flaws of the psychological decision-making mechanisms, social context and cultural influences. This is not to say that the actual content of fairness preferences is a random product of chaotic processes. In order to understand how context-dependent preferences are formed, one must take into account the interaction between social and psychological factors. This necessitates different kinds of theory.

III *Homo reciprocans* needs governance, but for different reasons

Let us now turn to the question of how prospects for cooperation are altered by the existence of fairness preferences. The picture is rather mixed and certainly more complicated than with selfish individuals only. In a population of people acting like *homo oeconomicus*, cooperation is possible if it is in the interest of the individual to cooperate, e.g., in repeated interactions under the restricted conditions put down in the so-called Folk Theorem (sufficient probability of repetition and sufficiently low discount rate, in particular).¹⁶ However, if cooperation is not incentive-compatible, the problem of opportunism may arise. In those situations, sanctions or some other mechanism have to be employed in order to make cooperation possible. Compared to that, it is at the same time easier and more difficult to sustain cooperation in situations with a considerable share of fair-minded people, as we shall see. This has extensively been studied with an experimental game called the Public Good Game, which is a variant of the Prisoner's Dilemma with multiple players (usually 4).¹⁷ Of course, the Public Good Game involves several (albeit few) actors, whereas cooperation in contractual relations is often bilateral. But seminal findings have been made with this research tool and it seems plausible that these findings will, in principle, carry over to interactions with fewer participants. In the Public Good Game, people receive an endowment that they can (in part or as a whole) keep to themselves or invest in a public good. In the latter case, because it is assumed that cooperation produces an added value, the sum invested in the public good is enlarged and distributed to all participants such that a contributor always gets back less from his own contribution than he invested, but on aggregate participants are better off. If a game of this kind is played only once, or for a known number of rounds, selfish individuals should always contribute nothing, i.e., cooperation should not occur.

16 E. Rasmusen, *Games and Information* (Malden, MA: Blackwell, 2nd ed. 1994), 124 et seq.

17 E. Fehr / S. Gächter, 'Cooperation and punishment in public goods experiments' *American Economic Review* 90 (2000), 980.

1 Reciprocity and its limits

Instead, people tend to invest 40 to 60 % of their endowment in one-shot Public Good Games.¹⁸ Thus, a considerable share of people does not behave selfishly right from the start. Their behaviour, however, is driven by the expectation that the others would also behave cooperatively.¹⁹ The reciprocal nature of their motivation to cooperate becomes apparent when the game is played for several rounds. These games also start with comparable contribution rates in the first round, but cooperation then declines inescapably until only about 10 % of participants still contribute.

a) *The problem of heterogeneous populations*

Judging from the result, it looks as if people were behaving selfishly in public-goods situations, as can be observed in many real-life situations. However, this is not because most participants are motivated by selfishness, but rather because a majority of fair-minded people responds to the opportunistic behaviour of a selfish minority by repaying in kind. Hence, the majority refuses to cooperate further for reasons of fairness. These actors can be described as “conditional cooperators”.²⁰ Their cooperation is conditioned on the cooperation of their other fellow actors. Although considerable caution is appropriate with regard to the external validity of such estimates, the proportion of different types of behaviour seems to be roughly around 30 % of participants acting selfishly, 60 % acting reciprocally and 10% acting altruistically, at least in some games, and according to some experimental tests.²¹

b) *Biased reciprocators*

However, although reciprocal behaviour in strategic games is often in fact motivated not by purely strategic considerations, but by concerns for fairness, this motivation should not be idealized either. In particular, people do not behave as impartially as should be expected from a truly ethical point of view. For example, conditional cooperators indeed reciprocate, but they systematically tend to contribute a bit less than they suppose others will do.²² What is showing here is perhaps one of the most pervasive characteristics of human motivation, the so-called self-serving bias. It is important to recognize that the self-serving bias does not result from conscious strategic considerations or from overtly selfish motives. Instead, self-

18 E. Fehr / B. Rockenbach, 'Human altruism: economic, neural, and evolutionary perspectives' *Current Opinion in Neurobiology* 14 (2004), 784, 785; U. Ones / L. Putterman, 'The ecology of collective action' *Journal of Economic Behavior and Organization* 62 (2007), 495, 496.

19 C. Engel / S. Kube / M. Kurschilgen, 'Can we manage first impressions in cooperation problems?' *Preprint* 2011/5 (Bonn: Max Planck Institute for Research on Collective Goods, 2011).

20 Fehr and Schmidt, n 10 above; E. Fehr / A. Falk, 'The psychological foundations of incentives' *European Economic Review* 46 (2002), 687, 692, 711.

21 U. Fischbacher / S. Gächter / E. Fehr, 'Are people conditionally cooperative?' *Economics Letters* 71 (2001), 397; U. Fischbacher / S. Gächter, 'Social preferences, beliefs, and the dynamics of free riding in Public Goods' *American Economic Review* 100 (2010), 541; B. Herrmann / C. Thöni, 'Measuring conditional cooperation' *Experimental Economics* 12 (2009), 97, R. Kurzban / D. Houser, 'Experiments investigating cooperative types in humans' *Proceedings of the National Academy of Science* 102 (2005), 1803; T. Page / L. Putterman / B. Unel, 'Voluntary association in public goods experiments' *Economic Journal* 115 (2005), 1032.

22 Fischbacher and Gächter, *ibid.*

servingly biased motivations can and often do appear to the actor to be objective, impartial and unbiased. This is so because the bias results from an unconscious distortion of implicit cognitive processes, particularly from distorted information search and information processing, which are not accessible to or usually are not accessed by introspection.²³ Even worse, while our own judgments appear authentic and objective, we are at the same time suspicious against others and expect their judgments to be biased.²⁴ Now, the decisive factor determining the strength of self-serving bias is ambiguity, be it of factual knowledge or of normative standards, because ambiguity creates the leeway in which information processing can be deflected in one or the other direction. Two survey studies provide a telling example of the unconscious nature of the self-serving bias. They were conducted with judges in the U.S. and in Switzerland, and who could be more impartial than judges? The judges were asked to estimate how well they perform compared to other judges, with regard to how often their judgments are overturned on appeal, and to rank themselves accordingly in one of four quartiles (best, second-best, etc.). Of course, these judges did not have the necessary statistical information at hand to make accurate estimates and were hence confronted with ambiguity. But with errors randomly distributed their self-evaluations should lead to the quartiles approximately being populated evenly. Instead, the upper two quartiles were heavily overpopulated: 88 % of U.S. judges and 92 % of Swiss judges rated themselves above average (the unbiased rating would have been 50 %) and 56 % of U.S. judges, but only 20 % of Swiss judges, saw themselves in the best quartile (25 % being the unbiased rating).²⁵ Unfortunately, but not unexpectedly, the self-serving bias also affects fairness judgments.²⁶

2 Sanctioning for the sake of fairness

On the one hand, one lesson from standard economic theory remains: cooperation cannot be sustained without institutional intervention, as would have been predicted under the assumption of selfishness. However, on the other hand, a better understanding of why cooperation is so fragile emerges. It is due mainly to the heterogeneity of people with regard to social preferences. Cooperation does not fail because everybody is selfish, but because the selfish behaviour of some (for whatever reason) drives fair-minded people also to behave selfishly (out of fairness concerns). Thus, selfish behaviour prevails in the end despite preferences for fairness, or rather because of preferences for fairness.

23 E. Pronin, 'Perception and misperception in human judgment' *Trends in Cognitive Science* 11 (2006), 37.

24 E. Pronin, 'How we see ourselves and how we see others' *Science* 320 (2008), 1177, 1178.

25 C. Guthrie / J. Rachlinski / A. Wistrich, 'Inside the judicial mind' *Cornell Law Review* 86 (2001), 777; M. Schweizer, *Kognitive Täuschungen vor Gericht* (Diss. iur., Univ. Zurich, 2005; available at <http://www.decisions.ch/dissertation.html>).

26 L. Babcock / G. Loewenstein, 'Explaining bargaining impasse: The role of self-serving biases' *Journal of Economic Perspectives*, 11 (1997), 109; J. Konow, 'The effects of information and stakes on fairness biases and dispersion' *Social Justice Research* 18 (2005), 349.

a) *The balancing function of sanctions*

One of the most interesting lessons to be learned from this with regard to governance concerns the role of sanctions. Insofar as behaviour is driven by reciprocity, the relationship between actors becomes paramount. This also changes the role of sanctions. Their objective is not confined anymore to correcting individual material incentives, i.e., to raise the expected costs of opportunistic behaviour to such a degree that it is in the own selfish interest of individuals to cooperate. This objective exhausts the task only with regard to the small fraction of individuals that behave selfishly. With regard to reciprocal actors, sanctions assume an additional objective of a quite different kind. Their behaviour is affected by sanctions not only and not so much as sanctions are addressed directly to them. They are also affected indirectly insofar as sanctions regulate the behaviour of other actors. This is so because reciprocal types condition their behaviour on the perceived fairness of the behaviour of others, repaying their fairness or unfairness in kind. Correcting material incentives with sanctions is therefore less necessary and sometimes less sufficient to keep up cooperation. It is less necessary because fair-minded people cooperate in order to reciprocate the cooperative behaviour of others. Here, social or moral incentives, as one might call them, provide a substantial part of the motivation to cooperation. However, if fair-minded people are faced with opportunistic behaviour on the part of their interaction partners, social or moral incentives work in the opposite direction and make cooperation even more difficult than with selfish individuals. The offense of being exploited unfairly renders such situations much worse than just losing money. Its behavioural effects, furthermore, are not constrained to silencing positive reciprocity. Being exploited regularly also triggers negative reciprocity, i.e., a strong urge to punish the malefactor, which exaggerates the destructive impact of defection on cooperation. In order to sustain cooperation among fair-minded people, sanctions thus also have to get social or moral incentives right. They have to assume the additional objective of preserving or restoring the balance in relationships between actors. Put more practically, the objective of sanctions is to sustain cooperation indirectly, that is, to preserve voluntary cooperation on the side of the majority of reciprocal actors by preventing opportunism on the side of selfish actors.

b) *Decentralised sanctioning, counter-punishment, and autonomy*

As we have seen, if *homo reciprocans* is involved, sanctioning of selfish behaviour becomes even more crucial. But institutions do not only have to be concerned with keeping up *homo reciprocans'* cooperative spirits by safeguarding him against exploitation. They also have to deal with *homo reciprocans'* powerful urge to punish opportunistic behaviour himself. Thus, fairness preferences in the form of negative reciprocity are themselves a source of sanctioning. Such decentralised punishment by the parties themselves can supplement or substitute centralised sanctioning administered by the state, but negative reciprocity also creates the need to tame private punishment. A series of public-goods experiments addresses this interaction of reciprocity, sanctioning and cooperation.

As said before, in Public Good Games without institutional intervention, cooperation declines in a vicious circle in which reciprocal actors adjust their contributions downward until a baseline of almost ubiquitous defection is reached. This negative dynamics can be reversed and almost complete cooperation achieved if experimental subjects are given the opportunity to punish defectors after each round the game was played, as has been shown in a famous series of experiments.²⁷ With regard to the selfish minority, decentralised punishment, if sufficiently serious, fulfils the function envisioned by standard economic theory. It eliminates the incentive to defect and causes them to cooperate. With regard to the reciprocal majority, decentralised punishment provided the necessary moral or social incentives for cooperation. It eliminated defection and induced cooperation on the part of fellow actors, thus creating the basis for positive reciprocity. For standard economic theory, the puzzling question was why people punish at all, despite punishment being costly. From a traditional point of view, a system of sanctioning is a higher-order public good in that it benefits everyone who benefits from the cooperative provision of the first-order public good.²⁸ But there is no shortage of voluntary participation in the sanctioning system, due to negative reciprocity.²⁹

However, voluntary decentralized punishment is not only good news. Because the urge to punish is so strong, an overprovision of sanctions can easily result from it. For example, the willingness to punish persists even when punishment becomes inefficient because the costs of sanctioning (borne by punisher and punishee) exceed gains of cooperation.³⁰ Moreover, it is not always co-operators punishing defectors, but also defectors punishing co-operators. Punishment is exerted not only for cooperative reasons, but also with an anti-cooperative intention. When given the opportunity to punish the punishers, punishment that was intended to induce cooperation is in turn being punished by defectors, leading to a decline in cooperation.³¹ In this regard, a series of cross-cultural experiments is quite illuminating.³² It showed

27 E. Ostrom / J. Walker / R. Gardner, 'Covenants with and without a sword' *American Political Science Review* 86 (1992), 404; E. Fehr / S. Gächter, 'Cooperation and punishment in Public Goods' *American Economic Review* 90 (2000), 980; E. Fehr / S. Gächter, 'Altruistic punishment in humans' *Nature* 415 (2002), 137; the same effect shows in centralized punishment: see R. O'Gorman / J. Henrich / M. Van Vugt, 'Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation' *Proceedings of the Royal Society B* 276 (2009), 323; D. Baldassarri / G. Grossman, 'Centralized sanctioning and legitimate authority promote cooperation in humans' *Proceedings of the National Academy of Science* 108 (2011), 11023.

28 E. Fehr / S. Gächter, 'Altruistic punishment in humans' *Nature* 415 (2002), 137.

29 Called "altruistic punishment"; see Fehr / Gächter, *ibid.*; for altruistic punishment by third parties, see, e.g., H. Bernhard / E. Fehr / U. Fischbacher, 'Group affiliation and altruistic norm enforcement' *American Economic Review* 96 (2006), p. 217; P. Lewisch / S. Ottone / F. Ponzano, 'Free-Riding on Altruistic Punishment?' *Review of Law and Economics* 7 (2011), 165; for an experiment on different possible motivations for punishment, see A. Niklisch / I. Wolff, 'Cooperation norms in multiple-stage punishment' *Journal of Public Economic Theory* 13 (2011), 791.

30 A. Botelho *et al.*, 'Social norms and social choice' *University of Central Florida Department of Economics Working Paper* (Orlando: University of Central Florida, 2005).

31 L. Denant-Boemont / D. Masclet / C. Noussair, 'Punishment, counterpunishment and sanction enforcement in a social dilemma experiment' *Economic Theory* 33 (2007), 145; N. Nikiforakis, 'Punishment and counter-punishment in Public Good Games' *Journal of Public Economics* 92 (2008), 91. For spiteful and strategic punishment, see also K. Jensen, 'Punishment and spite, the dark side of cooperation' *Philosophical Transactions of the Royal Society B* 365 (2010), 2635.

32 B. Herrmann / C. Thöni / S. Gächter, 'Antisocial punishment across societies' *Science* 319 (2008), 1362; see also S. Gächter / B. Herrmann, 'Reciprocity, culture and human cooperation' *Philosophical Transac-*

that anti-cooperative punishment exists in different cultures in different intensity. How prospects for cooperation are, crucially depends on the (culturally varying) amount of anti-social punishment. Because cooperative punishment is deterred or foiled by anti-cooperative punishment, cooperation is doomed to fail to the degree in which anti-social punishment is common to a culture. In this regard, it has proved telling to look at the relationship between punishing behaviour (as observed in the experiments) and social norms of civic cooperation (as elicited in the World Value Survey). Social norms of civic cooperation refer to people's attitudes towards tax evasion, abuse of the welfare state or dodging fares on public transport. These norms show a strong statistical correlation with cooperative behaviour: the stronger norms of civic cooperation are in a society, the harder the cooperative punishment in the experiments is. However, the weaker those norms are, the stronger anti-cooperative punishment is; that is, the harsher cooperators are being punished. It is not far-fetched to assume that strong norms of civic cooperation act as a constraint on antisocial punishment. Put more generally, the performance of formal institutions like contracts often depends on their interaction with social norms.³³

The existence of anti-cooperative punishment also reveals that being subjected to peer punishment is not always considered a preferable environment for cooperation. This shows in experimental games when people can choose either to play a game with or a game without the opportunity to punish.³⁴ In this case, the clear majority opts against punishment at first. However, after having experienced the devastating effects of missing sanctioning, almost everybody eventually switches to the game with punishment where now almost complete cooperation is achieved.³⁵ Here an aversion against sanctioning becomes visible, which probably stems from the fact that monitoring and coercion impairs one's autonomy.³⁶ That autonomy is a decisive factor is corroborated by research showing that cooperation increases substantially if a sanctioning mechanism is not imposed by fiat, but people can vote for or against it.³⁷ If chosen autonomously, participants apparently seem to find a sanctioning mechanism rather helpful than annoying. In a sense governance is faced with a dilemma regarding punishment: in order to sustain voluntary cooperation, opportunistic behaviour has to be suppressed by sanctions, but this interferes with people's autonomy and signals distrust. But at the same

tions of the Royal Society B 364 (2009), 791; S. Gächter / B. Herrmann / C. Thöni, Culture and cooperation, *Philosophical Transactions of the Royal Society B* 365 (2010), 2651.

- 33 S. Magen, 'Fairness, Eigennutz und die Rolle des Rechts' in C. Engel et al. (eds), *Recht und Verhalten* (Tübingen: Mohr Siebeck, 2007), 261; id., 'Zur Interaktion von Recht und sozialen Normen bei der dezentralen Bereitstellung von Gemeinschaftsgütern', in M. Stolleis and W. Streek (eds), *Aktuelle Fragen zu politischer und rechtlicher Steuerung im Kontext der Globalisierung* (Baden-Baden: Nomos, 2007), 185.
- 34 Ö. Gülerk / B. Irlenbusch / B. Rockenbach, 'The competitive advantage of sanctioning institutions' *Science* 312 (2006), 108.
- 35 Ö. Gülerk et al., *ibid.*; M. Sutter / S. Haigner / M. Kocher, 'Choosing the carrot or the stick?' *Review of Economic Studies* (2010), 77, 1540
- 36 For the valuation of autonomy see B. Frey / R. Jegen, 'Motivation crowding theory' *Journal of Economic Surveys* 15 (2001), 589; B. Frey / F. Oberholzer-Gee, 'The cost of price incentive' *American Economic Review* 87 (1999), 746.
- 37 S. Tontrup / W. Gaissmaier / H. Gong, 'The comparative advantage of legitimate procedures' (working paper, unpublished) (cooperation increased, e.g., from 58 to 85 % average contributions); see also D. Baldassarria and G. Grossman, n 27 above.

time, one has to maintain trust and a cooperative mindset that is impaired by control and coercion. Therefore it is crucial for the success of cooperation that sanctioning mechanisms are viewed as legitimate. Procedures might be the key to provide this legitimacy.

3 Coordinating fairness

There is another important issue that should be addressed here. It concerns the indeterminacy of fairness. Our previous conclusions about cooperation based on fairness have to be qualified in this regard. They are all premised on the presumption that people at least practically settle on a common understanding of what fairness means in a given context. In the Public Good Games used in the experiments reported above, selecting a fairness standard did not pose a serious problem. Since every participant was given the same endowment and benefited from the public good in the same way, equal contributions were the obvious standard of fair behaviour. But as soon as people differ in relevant respects (wealth, income, abilities, productivity, sensitivity, etc.) several standards compete. Actors then are faced with considerable leeway regarding the choice and operationalization of the criterion of justice. As always when faced with ambiguity, each party's choice of a standard is biased by self-interest. In real-life situations, which are usually fraught with differences and asymmetries, psychological indeterminacy of fairness is a permanent and serious source of conflict. It is also an important reason for institutional intervention. Considered in isolation, the selection of a standard of fairness seems to be a usual conflict over distributional questions. Viewed in the context of cooperation, this choice rather concerns a higher-order coordination problem: people are better off if they settle on a common standard because they can then reap the benefits of cooperation, but their preferences differ as to which standard should be chosen.

From a behavioural perspective, it seems to me that dealing with this higher-order coordination problem is one of the most important functions that institutions in general, and the law in particular, fulfil with regard to fairness and justice. But how, by which ways, can institutions in general, and the law in particular, shape and influence individual fairness judgments? One psychological mechanism, or rather a group of somewhat similar mechanisms, that seems of particular importance in this regard is framing. The term “framing” sometimes refers to a rather specific and sometimes to a more general effect. In its specific meaning, framing relates to prospect theory. Prospect theory posits that valuations of alternatives are not based on absolute outcomes, but depend on whether the outcome is framed as a gain or as a loss compared to a reference point, with losses weighing heavier than gains (loss aversion) and this difference leading to risk-seeking in the face of losses and risk aversion in the face of gains.³⁸ Changing the reference point by describing a decision problem in a different way can thus alter individual valuations by changing the frame from gains to losses or vice versa. Similar effects, it can be assumed, occur in fairness judgments. For example, not reciprocating a gift

38 D. Kahneman, 'A perspective on judgment and choice' *American Psychologist* 58 (2003), 697, 703 et seq.; for a documentation of seminal papers, see D. Kahneman and A. Tversky (eds), *Choices, Values and Frames* (Cambridge, UK: Cambridge Univ. Press, 2000).

seems to be less reprehensible than taking something of the same value from someone. Put more generally, in the context of fairness, reference points determine whether a given fairness question is a matter of positive reciprocity or of negative reciprocity.

In a second meaning, the term “reference point” does not refer to a turning point between losses and gains, but to social comparisons in general, e.g., to the questions of what the others get in a similar situation.³⁹ Social comparisons of this kind have a huge impact on fairness preferences. For example, if a proposer plays several Ultimatum Games with different responders simultaneously and responders know what the proposer offers to the other players, their willingness to accept an offer as fair crucially depends on the offers made to the other players.⁴⁰ But not only reference transactions serve as reference points. Familiar social norms or salient institutions can also serve this purpose. In this more general context of social comparisons and normative expectations, framing impacts on behaviour by providing (direct or indirect) information about reference transactions or about relevant social norms or institutions. These cues then influence the fairness standard that individuals adopt – not necessarily in a conscious way – to form their fairness judgment. For example, cooperation in a Prisoner's Dilemma experiment increased substantially only because the game was labelled “Community Game” instead of “Wall-Street-Game”.⁴¹

IV Fairness in Contract Governance

Contracts enable private parties to create obligations enforceable at law by mutual exchange of promises (under the conditions specified by the relevant law). A traditional Law and Economics point of view would consider such an institution to affect behaviour by altering selfish incentives through the expectation of enforcement, hence by sanctions. From a behavioural point of view, contracts might influence behaviour quite differently. Contracts obviously often serve as a means to institutionalize cooperation in form of mutually beneficial exchange. With *homo reciprocans* contracting, contracts will be accompanied by expectations of fairness.⁴² These fairness expectations might not be recognised by the law as relevant, but it is safe to assume that they influence the formation and execution of the contract by the parties and should hence be taken into account by Contract Governance. For example, a contract design proposed by economic contract theory as an efficient mechanism will be avoided if parties regard it as unjust. If chosen anyhow, it might have detrimental effects.⁴³ Put more generally,

39 G. Jasso, 'The theory of comparison processes', in: P. Burke (ed.), *Contemporary Social Psychological Theories* (Stanford: Stanford Univ. Press, 2006), 165; D. Kahneman / J. Knetsch / R. Thaler, 'Fairness and the assumptions of economics' *Journal of Business* 59 (2003), 697.

40 E.g., D. Alelwell / A. Nicklisch, 'Wage differentials and social comparison: An experimental study of interrelated ultimatum bargaining' *International Review of Law and Economics* 29 (2009), 210.

41 L. Ross / A. Ward, 'Naive realism in everyday life', in E. Reed et al. (eds), *Values and Knowledge* (Hillsdale, NJ.: Erlbaum, 1996), 103.

42 E. Fehr / S. Gächter / G. Kirchsteiger, 'Reciprocity as a contract enforcement device – experimental evidence' *Econometrica* 65 (1997), 833.

43 E. Fehr / A. Klein / K. M. Schmidt, 'Fairness, incentives and contractual incompleteness' *Working Paper No. 72* (Zurich: Institute for Empirical Research in Economics, Univ. of Zurich, 2001; see also E. Fehr /

fairness may support or impede a contract's execution, or it may interfere with it otherwise. These questions are of importance for any behavioural account of contract law, but they have only started to be addressed systematically. Thus, we have to confine ourselves to some examples that give an impression about the direction in which future research might proceed.

1 Statutes as reference points

In 2001, the German Distance Selling Act (now § 312b (1) BGB), implementing the EU Distance Selling Directive, introduced a mandatory withdrawal right for buyers in distant selling contracts. Since most distance sellers had already granted a withdrawal right on a voluntary basis, the revision of the law did not change the rights of buyers in effect, but only the motives of their provision. Nevertheless, the rate of returns increased from 24,2% to 35%.⁴⁴ For example, clothes were increasingly ordered in different sizes, with people sending back the unfitting ones. Georg Borges and Bernd Irlenbusch examined the reason for this change of behaviour with the help of an experiment.⁴⁵ Their results suggest that a voluntary provision of a withdrawal right is perceived as friendly such that buyers reciprocate by not exploiting sellers too heavily. However, if withdrawal rights are obligatory under the law, buyers may feel entitled to return goods, which eliminates the need to reciprocate. Also, sellers now lack the opportunity to show friendly intentions. As a consequence, Borges and Irlenbusch argue, the motivation to treat distance sellers fairly is crowded out by the law.⁴⁶ In my view, this interpretation does not exactly hit the point. The law presumably did not turn consumers into selfish beings, but changed the content of their fairness expectation such that they now felt morally entitled to return purchases. In their (maybe biased) perceptions, their reaction was not selfish, but righteous. So what the revision of the Distance Selling Act did in behavioural terms was to change the reference points relative to which consumers formed their fairness preferences.

This research is but an example of what has been called the “expressive function” of the law.⁴⁷ In standard Law and Economics, this function is confined to providing focal points in coordination problems,⁴⁸ but cannot contribute to the solution of problems of cooperation. With preferences for fairness, expressive aspects of the law assume a much more prominent and maybe crucial role.⁴⁹ They correspond to the need to coordinate individual standards of

A. Falk, 'Wage rigidity in a competitive incomplete contract market' *Journal of Political Economy* 107 (1999), 106.

44 G. Borges / B. Irlenbusch, 'Fairness crowded out by law: An experimental study on withdrawal rights' *Journal of Institutional and Theoretical Economics* 163 (2007), 84, 88; for a discussion of the phenomenon and its implications, see S. Bechtold *Die Grenzen zwingenden Vertragsrechts* (Tübingen: Mohr Siebeck, 2010), 110 et seq.

45 Borges and Irlenbusch, n 44 above.

46 Borges and Irlenbusch, n 44 above, 96 et seq.; for motivation crowding theory, see Frey and Jegen, n 36 above.

47 R. McAdams, 'An attitudinal theory of expressive law' *Oregon Law Review* 79 (2000), 339; on the meaning of the term 'expressive' in this regard see also L. A. Kornhauser, 'No best answer?' *University of Pennsylvania Law Review* 146 (1998), 1599, 1624 et seq.

48 R. McAdams, 'A focal point theory of expressive law' *Virginia Law Review* 86 (2000), 1649.

49 S. Magen, *Gerechtigkeit als Proprium des Rechts* (habilitation thesis, Univ. of Bonn, 2009), 183 et seq.

fairness if cooperation is to be sustained among reciprocal actors. The behavioural mechanism by which the law may achieve this is by providing salient and legitimate reference points and by framing interactions that take place in the shadow of the law. As the above example of the Distance Selling Act teaches us, the law has expressive effects anyhow, whether they are intended or not.⁵⁰ It also tells us that ignoring these effects may foil the purpose of the law or may inadvertently produce otherwise undesired results. Thus, Contract Governance seems to be well advised to take them into account.

2 Contracts as reference points

According to standard incomplete contracts theory, it can be advantageous for parties who want to enter into a long-term contract, but are faced with uncertainty about relevant aspects of the future state of the world, to form an incomplete contract first and leave the completion of contractual terms to future renegotiations when the actual state of the world has been revealed. If *homo oeconomicus*-like behaviour is assumed, these renegotiations will be conducted in a Coasian way leading to an (ex-post) efficient bargain.⁵¹ However, (ex-ante) inefficiencies can arise because of the hold-up problem. Under the prospect of future renegotiations, rational parties will underinvest during the contractual relationship because non-contractible sunk investments will be exploited by the other party in later renegotiations.⁵² It is an interesting question whether the hold-up problem is alleviated when parties care about fairness, possibly keeping them from exploiting commitments made by the other party. If, however, parties are also affected by self-serving biases and if they sometimes also punish spitefully or too much, fairness-related cooperation problems as those described above are to be expected in the context of long-term contracting, too.

That this is, in fact, the case has been shown recently in a very illuminating line of work,⁵³ which seems quite exemplary with regard to possible directions of future research. It empirically tests⁵⁴ a behavioural theory on renegotiations in incomplete contracts put forward by Oliver Hart and John Moore.⁵⁵ Instead of relation-specific investments, the theory considers performance. Performance, it assumes, can be perfunctory or consummate, but only perfunctory performance is verifiable before courts and hence contractible. Nevertheless, reciprocal actors will be willing to provide consummate performance as long as they feel they are getting

50 See, e.g., B. Depoorter / S. Tontrup, 'The legal endowment effect', Working Paper (San Francisco: Hastings College of Law, 2011); id., 'Contract entitlement', Working Paper (San Francisco: Hastings College of Law, 2011).

51 On incomplete contracts, see P. Bolton / M. Dewatripont, *Contract Theory* (Cambridge, Mass.: MIT Press, 2005), 487 et seq.; O. Hart, 'Incomplete contracts' in S. N. Durlauf and L. E. Blume (eds), *The New Palgrave Dictionary of Economics* (2nd ed., Palgrave Macmillan, 2008).

52 See Bolton and Dewatripont, *ibid.*, 560 et seq.; S. Shavell, 'Contractual holdup and legal intervention' *Journal of Legal Studies* 36 (2007), 325.

53 O. Hart / J. Moore, 'Contracts as reference points' *Quarterly Journal of Economics* 73 (2008), 1; E. Fehr / O. Hart / C. Zehnder, 'Contracts, reference points, and competition – behavioral effects of the fundamental transformation' *Journal of the European Economic Association* 7 (2009), 561; id., 'Contracts as reference points — experimental evidence' *American Economic Review* 101 (2011), 493.

54 Fehr, Hart and Zehnder (2009); id. (2011); n 53 above.

55 Hart and Moore, n 53 above.

what they are entitled to. If, on the other hand, parties feel treated unfairly, they will withhold some part of consummate performance, dubbed 'shading' by the authors. Since the self-serving distortion of fairness judgments increases with ambiguity, shading should increase with ambiguity because parties will more often feel deprived of what they are entitled to. Against the backdrop of these considerations, long-term contracts assume quite a different function of which standard contract theory is not aware: contracts, according to Hart and Moore, serve as reference points determining a party's sense of entitlement.⁵⁶ If a contract is rigid, pins down future outcomes very precisely and leaves little room for disagreement and aggrievement, parties will rather not shade on their obligations as long as they get what is specified in the contract. If, however, a contract is more flexible and allows for more than one outcome, feelings of entitlement will be exposed to self-serving distortions. Then, Hart and Moore assume, each party will feel entitled to the outcome most favourable to it, whereby fairness-expectations will divert and shading increase. Hence, contracting parties are faced with a trade-off between rigidity (which reduces shading, but excludes adjustments to unforeseen contingencies) and flexibility (which allows for adjustments, but increases shading).

Experiments conducted by Ernst Fehr, Oliver Hart and Christian Zehnder using a two-stage game support this theory.⁵⁷ In the game, participants are uncertain at the time of contract formation whether production costs will be low (good state of nature) or high (bad state of nature) at the time of performance. At a first stage, buyers can choose between a rigid and a flexible contract. Prices are then negotiated in an auction in which sellers compete for contracts. In rigid contracts, the negotiated price is fixed, irrespective of the state of nature. In flexible contracts, the negotiated price forms the lower bound of a price range (the upper bound being given exogenously) from which the buyer can choose the actual price at the second stage. After contracting, the state of nature is revealed and buyers can choose a price if the contract is flexible. When costs turn out to be high, rigid contracts are not performed because they are not mutually beneficial anymore. Flexible contracts, in contrast, allow the buyer to choose a higher price and perform the contract. After the buyer has chosen the price from within the price range, the seller can choose between consummate performance (high quality) and perfunctory performance (low quality). Perfunctory performance reduces the buyer's payoffs substantially,⁵⁸ but also burdens the seller with a slightly higher (not lower!) cost.⁵⁹ Thus, the seller has no material incentive to shade because he has to pay a price for it. The second stage thus bears a certain similarity with an Ultimatum Game.⁶⁰

The results are very much in line with Hart and Moore's hypothesis of 'contracts as reference points'. First, both with rigid and flexible contracts competitive forces compel sellers to accept low prices (in rigid contracts) or lower price bounds (in flexible contracts) and prices and

56 Hart and Moore, n 53 above, 2 et seq. For contracts and loss frames, see R. R. W. Brooks / A. Stremitzer / S. W. Tontrup, 'Framing contracts: Why loss framing increases effort' *John M. Olin Centre for Studies in Law, Economics, and Public Policy Research Paper* No. 438 (New Haven, CT: Yale Law School, 2011).

57 Fehr, Hart and Zehnder (2009), 561, 568 et seq.; id. (2011), 493, 505 et seq.; n 53 above.

58 100 instead of 140.

59 25 instead of 20.

60 Further discussed by Fehr, Hart and Zehnder (2011), n 53 above, 519.

lower price bounds converge downwards to the competitive price. Competition hence leads to a pronouncedly uneven distribution of payoffs. With rigid contracts, sellers on average earn only a quarter of what buyers earn. With flexible contracts, actual prices chosen from the price range are somewhat higher, but sellers still receive only 40 % of buyers' payoffs on average.⁶¹ The question then is: how do sellers react to this? Selfish sellers should not punish at all, because punishment is costly (and hence not subgame perfect). Reciprocal sellers should shade when they perceive the offer as unfair. Since prices are lower in rigid contracts than in flexible contracts, inequity aversion should lead to more shading in rigid contracts. According to the 'contract as reference points' hypothesis, however, it should be the other way around because rigid contracts are unambiguous and thus serve better as reference points, whereas flexible contracts provide leeway for self-serving distortions, which increases perceived unfairness. Experiments confirm this hypothesis. Although sellers earn less from rigid contracts, they nevertheless still almost always provide consummate performance (94 % of the cases), whereas consummate performance drops significantly in flexible contracts (to 75 % of the cases).⁶² However, if contracts are not formed voluntarily by the parties, but terms and parties to the contract are just given exogenously, shading reappears in rigid contracts (consummate performance drops to 84 %) and the difference to flexible contracts is no longer statistically significant.⁶³ Somehow, rigid contracts seem to lose their legitimacy as reference points under these circumstances. Fehr, Hart und Zehnder take this observation as evidence that competition legitimizes the terms of a contract. The reason they suppose is that buyers can hide their unfairness behind a 'veil of competition' because outcomes are attributed rather to market processes.⁶⁴ Although there is something to be said for this explanation, it somehow misses the important point. If parties are matched and given the terms by fiat, there is no contract anymore at all, but rather a forced relationship. Not competition is missing here, but mutual agreement and consent. Thus, consent makes reference points legitimate and marks the crucial difference between rigid and flexible contracts: in rigid contracts, the seller consents to exactly the price he is later given. Therefore he accepts it as fair. In flexible contracts, the seller does not consent to any price, but to a lower price bound and the right of the buyer to determine the price later. Therefore the reference point in flexible contracts is not only more ambiguous (which it is, indeed), but the contract is referred to in a different context in which it assumes a different meaning. Flexible contracts do not serve as a reference point for performance based on a set price, but as a reference frame for deferred price negotiations. This not only entails more ambiguity, but also diminished legitimacy because the seller has not committed himself to anything more than a lower bound for negotiations yet to come. Apart from this, there is nothing specific he feels bound to later.

61 See Fehr, Hart and Zehnder (2011), n 53 above, 506.

62 Fehr, Hart and Zehnder (2011), n 53 above, 506 et seq. An analysis of individual behaviour confirms this result, see *ibid.*, 512 et seq.

63 Fehr, Hart and Zehnder (2011), n 53 above, 516 et seq.

64 *ibid.*, 497, 521.

3 Commitment through promises and legal procedures

This line of argument suggests, of course, that there might be other specific motivational forces at work in contractual relationship beyond reciprocity. Although this has not been researched in depth, there is indeed some initial evidence that the very fact of making a promise as well as the “procedure” of doing so with reference to the law may have a behavioural effect. With standard preferences, neither the exchange of promises nor reference to the law as such should make any difference, because without the backing of sanctions, such communication remains “cheap talk”. In experimental games, however, the mere exchange of promises usually enhances cooperation considerably. However, it is unclear whether this is due to fairness (not letting down others' legitimate payoff expectations) or because people have an intrinsic motivation to keep their word. At least one experiment suggests that the latter is the case.⁶⁵ In another experiment, it is shown that cooperation is further enhanced when an agreement is framed as a valid contract (citing legal norms) compared to a mere promise, although enforcement is possible in neither of the cases. Hence the procedure of concluding a contract with reference to the law seems to evoke an intrinsic motivation to comply beyond the moral commitment of a promise.⁶⁶

V Conclusion

The last two decades have witnessed tremendous progress towards a behavioural account of fairness and reciprocity, brought about by a combination of economic theories and empirical methods that is nowadays known as behavioural economics. The point of this research was not just to show what anyone except economists knew anyhow, that is, that people do care for justice and fairness. The point was to examine if and how individual concerns for fairness play out in strategic interactions and thus have to be regarded as a class of preferences of practical relevance in economic and other policy-related settings. In this regard, behavioural economics was successful in showing that fairness preferences influence behaviour in a way that renders predictions based on the selfishness assumption often, but not always, misleading. However, to know that man is rather a *homo reciprocans* than a *homo oeconomicus* does not mean that his behaviour can be foreseen better, because *homo reciprocans* is a rather impressionable being and his behaviour heavily depends on the specifics of the situation and the social context in which he is acting. This characteristic of fairness motives also casts doubt on the external validity of experimental results, not so much with regard to the general behavioural tendencies and dispositions detected, but with regard to the question how these general behavioural mechanisms will play themselves out in real life. What is needed here is a combi-

65 C. Vanberg, 'Why do people keep their promises?' *Econometrica* 76 (2008), 1467; see also M. Belot / V. Bhaskar / J. van de Ven, 'Promises and cooperation: Evidence from a TV game show' *Journal of Economic Behavior and Organization* 73 (2010), 396; T. Baumgartner / U. Fischbacher / A. Feierabend / K. Lutz / E. Fehr, 'The neural circuitry of a broken promise' *Neuron* 64 (2009), 756; G. Charness / M. Dufwenberg, 'Bare promises: An experiment' *Economic Letters* 107 (2010), 281.

66 S. Tontrup / B. Irlenbusch / M. Kurschilgen, *The expressive function of contracts* (working paper, unpublished); but see C. Vanberg, 'Voting on a sharing norm in a dictator game' *Journal of Economic Psychology* 31 (2010), 285.

nation of theory and experiments, corroborated by real-life data from field experiments or econometrics.⁶⁷ Thus, unfortunately, what is a clear progress in the scientific account of human social behaviour turns out to be a nuisance for the quest for solutions to specific governance problems because with fairness these now easily hinge on unexpected details. Whereas traditional rational choice theory often can come up with predictions that are robust but inaccurate, predictions based on fairness preferences are more realistic, but are, at the same time, notoriously incomplete and rather fuzzy. Nevertheless, quite a lot can be said about certain generic properties of fairness and reciprocity as driving forces of behaviour.

In this article, I tried to sketch some important features of these general working mechanisms of fairness based on a selective review of a small part of the literature. To understand those working mechanisms, it is crucial to distinguish mental, behavioural and social aspects of fairness and to see how they work together and what their respective contributions are in sustaining cooperation or avoiding opportunistic behaviour. Experimental game theory has focused mainly on the behavioural level and developed two types of models of preferences for fairness: inequity aversion, concerned with the distribution of outcomes, and reciprocity, concerned with the fairness of intentions. Both assume preferences for fairness to interact with selfish preferences that remain a primary motivational force. Fairness judgments, however, also allow for several proportional criteria of distribution, as Equity Theory has shown. The (psychological) necessity to select among different criteria of fairness and their operationalization renders fairness judgments indeterminate and makes them receptive for social and contextual influences.

Compared to *homo oeconomicus*, cooperation is both easier and more difficult with *homo reciprocans*. It is easier because reciprocity repays cooperation with cooperation. It is more difficult because opportunistic behaviour by selfish types, usually mixed into real-life populations, triggers negative reciprocity on the side of reciprocal actors causing a downward spiral that eventually extinguishes cooperation. Furthermore, even fairness judgments are unconsciously biased toward self-interest if and insofar as factual knowledge or normative standards are ambiguous. Therefore, sanctions remain crucial to sustain cooperation, but they assume a quite different objective compared to *homo oeconomicus*. With *homo reciprocans*, sanctions have to provide social or moral incentives in order to sustain positive reciprocity, and they do so indirectly by preventing opportunism and thus keeping social relations fair or balanced. Negative reciprocity provides sufficient motivation to leave punishment to the parties, but excess punishment and anti-cooperative punishment, combined with self-serving perceptions, call for institutional intervention in order to prevent a vicious circle. Although people in general show an aversion against punishment institutions, they are accepted as legitimate if they are chosen autonomously.

67 For a commendable example, see B. Frey / B. Torgeler, 'Tax morale and conditional cooperation' *Journal of Comparative Economics* 35 (2007), 136.

Indeterminacy of fairness creates the need to settle on shared fairness standards, if cooperation is to be based on positive reciprocity and protected against negative reciprocity. Since fairness judgments intrinsically rely on reference points and reference transactions, institutions can contribute to this higher-order coordination problem by providing those references. In Contract Governance, both statutes and contracts can serve this purpose through their expressive effect. For example, making a withdrawal right obligatory by law – it had hitherto been granted by distance sellers on a voluntary basis – can increase returns because the revision of the law also shifted the framing of withdrawal from a friendly concession to an action the seller is obliged to endure. In long-term contractual relations, initial contracts can serve as reference points for subsequent renegotiations. This can reduce perfunctory performance or other forms of opportunistic behaviour if contracts are unambiguous rather than flexible. Finally, the mere fact of making a promise and the 'procedure' of doing so with reference to the law induces a commitment to the contract beyond reciprocity. In short, with *homo reciprocans* as a party to the contract, contractual relationships are necessarily accompanied by expectations of fairness and reciprocity. Contract Governance is thus charged with the task of institutionalising and taming sanctions and of coordinating fairness standards by managing frames and reference points via the expressive dimension of statutes and contracts.