

2016

Order Ethics and the Problem of Social Glue

Christoph Luetge

Bluebook Citation

Christoph Luetge, *Order Ethics and the Problem of Social Glue*, 12 U. St. Thomas L.J. 271 (2016).

This Article is brought to you for free and open access by UST Research Online and the University of St. Thomas Law Journal. For more information, please contact lawjournal@stthomas.edu.

ARTICLE

ORDER ETHICS AND THE PROBLEM OF SOCIAL GLUE

CHRISTOPH LUETGE*

I. THE PROBLEM SETTING

It is an old question of philosophy—from Plato and Aristotle to Hobbes, Hume, Kant, and Hegel—to ask what holds human societies together. However, during the last century, and especially since the dawn of globalization, this question has gained much importance as societies face a previously unknown degree of cultural, social, and economic pluralism.¹ It is not clear how or even whether age-old answers to this problem can have bearing on modern social problems. To give only one example of how modern problems differ from past ones, consider how collective structures and social arrangements in modern societies are far more vulnerable to the actions of individuals or small groups than in earlier days, which has been known in theory long before, and has been made dramatically clear by the events on September 11, 2001.

Many contemporary positions in social and political philosophy (some of which will be discussed here) have been asking the following question: Granted that actors in modern societies strive for advantages and benefits, do we still need some kind of a *social glue* beyond this mere quest for advantages and benefits? I would like to call this kind of social glue a *moral surplus*.² A moral surplus is a moral capacity or capability that the citizens of a modern society have to adopt in order to keep their society stable. This

* Christoph Luetge received his Ph.D. in philosophy in 1999 from the Technical University of Braunschweig and his habilitation from the University of Munich in 2005. He worked as a visiting researcher in Pittsburgh, San Diego, and Venice. Since 2010, he holds the Peter Löscher Chair of Business Ethics at Technical University of Munich. The author thanks Wulf Kaal and the University of St. Thomas School of Law for the opportunity to present some of these ideas at the Symposium on Global Business Ethics.

1. Compare, for example, JOSEPH E. STIGLITZ, *GLOBALIZATION AND ITS DISCONTENTS* (2002) and JAGDISH BHAGWATI, *IN DEFENSE OF GLOBALIZATION* (2004), for accounts of the phenomenon of globalization.

2. For more on this term, see generally CHRISTOPH LUETGE, *ORDER ETHICS OR MORAL SURPLUS: WHAT HOLDS A SOCIETY TOGETHER?* (2014).

moral surplus, according to these positions, is necessary in addition to rules and incentive structures. It may, however, differ in degree and strength.

In this paper, I will discuss some contemporary philosophical positions, which (except for the last one—K. Binmore) argue that modern societies do indeed require a moral surplus to prevent human societies from destabilizing. The positions analyzed in the following sections are J. Habermas's discourse ethics, J. Rawls's political liberalism, D. Gauthier's contractarian moral theory, and finally, K. Binmore's game theory-based contractarianism. As I can dedicate only a short part of this article to each of these positions, I will restrict myself to the following question: Can the moral surplus in question remain stable in the face of opposing incentives?

This question can be explicated in the following, rather simple, way: The prime situational model for interactions that concern drastically disadvantageous incentives to all actors is the prisoners' dilemma (PD).³ In PD situations, actors behaving cooperatively are permanently faced with the possibility of being "exploited" by others, therefore they pre-emptively stop cooperating. This leads to a situation where rational, self-interested actors end up with a result that leaves everyone worse off and no one better off.⁴ Two of the four approaches discussed here (Gauthier and Binmore) regard the PD as their conceptual starting point.

I will argue that the question whether the moral surplus can remain stable must be answered negatively in each of the cases discussed, with the exception of Binmore. None of the first three surpluses can remain stable against opposing incentives in PD situations. I will start with Habermas's discourse ethics.

II. HABERMAS: RATIONAL MOTIVATION

In arguing for the importance of discourses as grounds for normative theory, Habermas makes a fundamental assumption—the participants of a discourse must allow their behavior, at least partially, to be motivated by a *rational motivation*.⁵ With the help of this concept, Habermas explicates the

3. See ROBERT AXELROD, *THE EVOLUTION OF COOPERATION* (1984), for conceptualizations of the prisoners' dilemma. Sometimes this situation is also called the *prisoner's* dilemma. I deliberately use the plural because it is an *interaction* problem for *both* actors, not just a unilateral decision problem for a single actor.

4. *Id.* PD situations means introducing sanctions, i.e., incentives. If this is done by way of institutions, then this is in agreement with my approach (*see infra* Part VI). If not, then it is unclear what could be meant by iterating a PD situation "in real life" *without* relying on institutions for stability.

5. See JÜRGEN HABERMAS, 2 *THE THEORY OF COMMUNICATIVE ACTION*, 26, 29 (Thomas McCarthy trans., Beacon Press) (1981) [hereinafter HABERMAS, *THEORY OF COMMUNICATIVE ACTION*]; JÜRGEN HABERMAS, *MORAL CONSCIOUSNESS AND COMMUNICATIVE ACTION* 58, 109 (Christian Lenhardt & Shierry Weber Nichol森 trans., MIT Press) (1990) [hereinafter HABERMAS, *MORAL CONSCIOUSNESS*]; JÜRGEN HABERMAS, *BETWEEN FACTS AND NORMS: CONTRIBUTIONS TO A DISCOURSE THEORY OF LAW* 5 (William Rehg trans., MIT Press) (1992) [hereinafter HABERMAS, *BETWEEN FACTS AND NORMS*].

difference between his two opposing types of actions: strategic and communicative. This difference represents a central antagonism in Habermas's theory. While strategic action (*strategisches Handeln*) is affected by incentives and sanctions, communicative action (*kommunikatives Handeln*) is led by rational motivation.⁶ But what, then, *is* rational motivation and where does it come from?

It is difficult to find an answer to this question in Habermas's writings. In *Theory of Communicative Action*, he simply postulates: "If arguments are valid, then insight into the internal conditions of their validity can have a rationally motivating force."⁷ Yet a few pages earlier, there is a footnote in which Habermas admits that the concept of a rational motivation has not yet been analyzed satisfactorily.⁸

In later works, the situation does not get much better. In *Diskursethik: Notizen zu einem Begründungsprogramm*, the necessity of a rational motivation is stressed several times,⁹ but only in one instance is a justification given. Here, Habermas relies on the illocutionary effects of a speech act.¹⁰ He assumes that speech acts can force actors to perform certain actions and refrain from others. The core of this assumption is the idea of a performative contradiction.¹¹ According to Habermas, a rationally motivated actor is led by the desire to avoid a performative contradiction.

It is doubtful whether a speech act has such a binding force. It seems that Habermas recognizes this problem himself, as he invents a dialogue with a fictitious skeptic who doubts exactly this binding force of speech acts.¹² However, Habermas responds that the skeptic may well stick to his position but then has to "be silent." He cannot escape the *Lebenswelt*, which, according to Habermas, is formed by cultural tradition and socialization, which in turn work through rational motivation. If he tried to escape, he would end in "schizophrenia and suicide."¹³

In the 1990s, Habermas's thought developed in new directions. He faced the problem of the skeptic again in "Remarks on Discourse Ethics." Here, Habermas first seems to weaken the power of rational motivation by attributing to moral norms only "the weak motivating force of good reasons."¹⁴ He goes as far as stating that "the validity of moral commands is subject to the condition that they are *universally* adhered to as the basis for

6. See HABERMAS, MORAL CONSCIOUSNESS, *supra* note 5, at 58.

7. HABERMAS, THEORY OF COMMUNICATIVE ACTION, *supra* note 5, at 29; *see also id.* at 42.

8. *Id.* at 26, 411 n.28.

9. See HABERMAS, MORAL CONSCIOUSNESS, *supra* note 5, at 58, 72, 109.

10. *Id.* at 58.

11. *See id.* at 80–81 (explicating further the idea of a performative contradiction).

12. *See generally* HABERMAS, MORAL CONSCIOUSNESS, *supra* note 5.

13. *Id.* at 102.

14. JÜRGEN HABERMAS, JUSTIFICATION AND APPLICATION: REMARKS ON DISCOURSE ETHICS 33 (Ciaran P. Cronin trans., MIT Press) (1993).

a general practice,”¹⁵ thus implying that individuals might be allowed to behave “immorally” when faced with possible exploitation by others. This would be the case in prisoners’ dilemma situations.¹⁶ However, in the rest of his article, Habermas does not consider this a problem of ethics. Rather, he writes, it can only be argued within the discourse of *law* that some norms might be valid yet not reasonable (*zumutbar*) because of their lack of general acceptance.

Consequently, this leads Habermas to include institutions in his conception, which were rather neglected in his earlier work. He recognizes that the problem of compliance to norms cannot be solved by postulates of rational motivation or by citing developmental psychology.¹⁷ In *Between Facts and Norms*, he explicitly develops an account of law and institutions. However, even here it is clear that he does not rely on sanctions and incentives alone as a means for governing modern societies. Rather, the citizens must still have certain characteristics—they have to accept not only the legal rules, but also normative claims resulting in turn from the idealized discourse assumptions.¹⁸ Habermas still assumes a “coordination of plans of actions”¹⁹ by language—thus still assuming the existence of rational motivation working via speech acts. Consequently, Habermas’s main claims are still intact in his second, institutional, phase.

Now the question mentioned above becomes relevant: Can rational motivation remain stable in view of opposing incentives? One problem with Habermas’s argument concerning speech acts is his reconstruction of the *Lebenswelt*.²⁰ In my view, there are alternative and less harmonious reconstructions of the *Lebenswelt* possible, in which the binding force of speech acts is much weaker. One major example is the game-theoretic approach that reconstructs all human interactions as “riddled” with dilemma situations like the prisoners’ dilemma.²¹ These situations can be either manifest (as in open market interactions with competition being obvious) or hidden. Hidden PD situations that have been overcome are exemplified prominently in all the institutions like police and jurisdiction that come into effect as a consequence of the social contract which enables the actors to escape from

15. *Id.* at 34.

16. For more on PD situations, see generally KEN BINMORE, *GAME THEORY AND THE SOCIAL CONTRACT: VOL. 1: PLAYING FAIR* (1994) [hereinafter BINMORE, *GAME THEORY VOL. 1*]; KEN BINMORE, *GAME THEORY AND THE SOCIAL CONTRACT: VOL. 2: JUST PLAYING* (1998) [hereinafter BINMORE, *GAME THEORY VOL. 2*]; KARL HOMANN, *VORTEILE UND ANREIZE: ZUR GRUNDLEGUNG EINER ETHIK DER ZUKUNFT* (Christoph Luetge ed., 2002) (in which each assign the PD a central role in their ethical or ethically relevant approaches).

17. In some (earlier) works, he makes extensive use of L. Kohlberg’s developmental psychology. See, e.g., HABERMAS, *MORAL CONSCIOUSNESS*, *supra* note 5.

18. See HABERMAS, *BETWEEN FACTS AND NORMS*, *supra* note 5, at 17–18, 459–62.

19. *Id.* at 34.

20. HABERMAS, *BETWEEN FACTS AND NORMS*, *supra* note 5.

21. See generally BINMORE, *GAME THEORY VOL. 1*, *supra* note 16; BINMORE, *GAME THEORY VOL. 2*, *supra* note 16.

the natural state (i.e., a PD situation). Sanctioning institutions are often solutions to prisoners' dilemma situations.

In a phenomenalist perspective that does not look beyond the surface, it appears the individuals who act according to these institutions have been moved by a rational motivation. However, in another perspective, the "deeper" structures behind become visible, and it becomes clear that it is not just rational motivation, but rather incentives and possible sanctions (in this case, informal) that hold this social practice in place. This casts doubt on Habermas's claim to have found the only reconstruction of everyday practices.

Moreover, while Habermas introduces institutions into his conception in later works, thus allowing for sanctioning by legal rules, he still relies on his original idea of rationally motivated individuals. Institutions are assigned only a secondary place, while the original moral surplus survives intact. Therefore, Habermas's conceptions can be regarded as a chief example of a moral surplus theory.

III. RAWLS: THE SENSE OF JUSTICE

Rawls's conception of justice as fairness (JF)²² is known so well, so I can limit myself to some very general remarks. I will focus here on the conception as it is presented in *Political Liberalism*.²³ Its main starting point is the idea that JF is political, not a metaphysical conception. For Rawls, this is the only conception of justice suitable for modern, pluralistic societies, where pluralism of values and norms must be regarded as a permanent condition.²⁴ The constitution of a modern state cannot be built on any comprehensive (e.g., philosophical or religious) doctrine, but only on an idea that avoids conflicts between these doctrines—a political conception. It can be characterized by the following four points:

(1) The conception must be freestanding. It must be possible to present this conception without reference to comprehensive doctrines. It should, however, be possible to find arguments for this conception from the point of view of different doctrines.²⁵ Rawls hopes that JF will eventually be supported by an overlapping consensus of reasonable doctrines.²⁶

22. See generally JOHN RAWLS, *A THEORY OF JUSTICE* (1971); JOHN RAWLS, *POLITICAL LIBERALISM* 3–40 (1993) [hereinafter RAWLS, *POLITICAL LIBERALISM*].

23. There are some interesting differences between the "Theory of Justice" and "Political Liberalism," but I cannot go into detail here.

24. See John Rawls, *Justice as Fairness: Political not Metaphysical*, 13 *PHIL. & PUB. AFF.* 223, 223–51 (1985) (discussing this concept for the first time).

25. See RAWLS, *POLITICAL LIBERALISM*, *supra* note 22, at 10–11.

26. *Id.* at 58–66.

(2) The conception will judge only the basic structure of a society, that is, only the basic institutions—like freedom of speech, property rules, etc.—which have extraordinary importance for the citizens.²⁷

(3) The main content of this conception are the two famous principles of justice. First, all people have equal claims “to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all.”²⁸ Second, social and economic inequalities are justified only insofar as they are “attached to positions and offices open to all under conditions of fair equality of opportunity” and they must be to the greatest advantage of the least advantaged.²⁹

(4) The acceptance of these two principles can be reconstructed as a choice of rational actors in an original position.³⁰ While the construction of this original position is less relevant here and has been discussed elsewhere,³¹ it is important to note that it is only a means of presentation, a think tool not to be understood in an ontological sense.³²

While the conception of justice described so far is generally known, it is only part of the story. Just like Habermas, Rawls regards a moral surplus as indispensable for the stability of modern societies. According to Rawls, the citizens of a modern state must have two “moral” capabilities. First, they must have personal conception of the good, and second, be motivated by a sense of justice.³³ While the first is rather unproblematic and will not be discussed here, the second is quite interesting and poses a number of problems.

Rawls characterizes the sense of justice³⁴ as an integral part of a well-ordered society³⁵ and defines it as “the capacity to understand, to apply, and normally to be moved by an effective desire to act *from (and not merely in accordance with)* the principles of justice.”³⁶ The central passage “to act from” is important—Rawls explicitly does *not* want to say that individuals comply with rules because of expected benefits. Rather (in my reading of

27. *Id.* at 11–12.

28. Here again, I use the wording employed by Rawls. RAWLS, *POLITICAL LIBERALISM*, *supra* note 22, at 5.

29. *Id.* at 6.

30. *Id.* at 22–28.

31. See Ronald Dworkin, *The Original Position*, in *READING RAWLS* 16, 16–52 (Norman Daniels ed., 1975).

32. See generally John Rawls, *Reply to Habermas*, 92 *J. PHIL.* 132 (Mar. 1995).

33. RAWLS, *POLITICAL LIBERALISM*, *supra* note 22, at 35.

34. The sense of justice concept appears first in “The Sense of Justice.” John Rawls, *The Sense of Justice*, 72 *PHIL. REV.* 281, 281–303 (1963).

35. A society is well-ordered according to Rawls if (a) the two principles of justice are generally accepted, (b) the basic structure conforms with these principles, and (c) the citizens possess a sense of justice. See RAWLS, *POLITICAL LIBERALISM*, *supra* note 22, at 35–39.

36. I use the more elaborated wording employed by Rawls rather than his shorter version. Compare RAWLS, *POLITICAL LIBERALISM*, *supra* note 22, at 302, with RAWLS, *POLITICAL LIBERALISM*, *supra* note 22, at 19.

this passage), the principles of justice must be complied with even if no personal benefits can be made out.

This becomes clear again in a second, more general definition of the sense of justice, which is defined here as the capability “to honor fair terms of cooperation.”³⁷ The phrase “to honor” is important—fair rules can be honored by all actors regardless of the incentives in effect. Rawls does not talk about governance by incentives in this context. The rules of a fair society are characterized as fair, but he does not say if they are to be sanctioned. In particular, he does not say how counterproductive consequences—in PD situations, for example—might be avoided. If my reading of Rawls is correct, he would have to call for compliance with norms or rules even if there were PD-like incentives running against it.³⁸

This is one of the main themes in JF, which becomes visible in several opposing concepts that Rawls constructs: the reasonable versus the rational, the *modus vivendi* versus the overlapping consensus, and the constitutional consensus versus the overlapping consensus.³⁹ In all three cases, Rawls’s moral surplus, the sense of justice, plays a decisive role. First, the individuals who are just rational, but not reasonable, lack the sense of justice. Second, an overlapping consensus can only be formed by individuals acting *from* the sense of justice. I will focus here on the second point.⁴⁰

Rawls emphasizes that JF has to be supported by an overlapping consensus of reasonable comprehensive doctrines and not just by a “mere” *modus vivendi*. He defines the *modus vivendi* as a contract between two parties with opposing goals. Both parties are interested in keeping the contract.⁴¹ However, they would break it if the situational conditions changed and their own interest could be better pursued at the others’ expense. According to Rawls, this situation cannot serve as a basis for social stability. A consensus founded only “on self- or group interests,” being only the result of “political bargaining,”⁴² must be regarded as inherently unstable—if the conditions changed, so would the *modus vivendi*, with the result of social stability eroding.⁴³ Rawls also assumes that a society collapses if its conception of justice and its democratic order are not “supported by at least a substantial majority of its politically active citizens.”⁴⁴

37. RAWLS, POLITICAL LIBERALISM, *supra* note 22, at 302.

38. *Id.* at 35 (“[C]itizens have a normally effective sense of justice *and so* they generally comply with society’s basic institutions.”) (emphasis added).

39. For the following, compare with LUTGE, *supra* note 2.

40. The difference between the rational and the reasonable has been widely analyzed. See, e.g., VICTORIA DAVION & CLARK WOLF, THE IDEA OF A POLITICAL LIBERALISM: ESSAYS ON RAWLS (2000); CHANDRAN KUKATHAS & PHILIP PETTIT, RAWLS: A THEORY OF JUSTICE AND ITS CRITICS (1990).

41. A *modus vivendi* is thus not comparable to a “cease fire” as opposed to a peace treaty, but is rather a “simple” (exchange) contract.

42. RAWLS, POLITICAL LIBERALISM, *supra* note 22, at 147.

43. *Id.*

44. *Id.* at 38.

In view of this, Rawls wants to build social stability on a stronger basis—on an overlapping consensus of reasonable comprehensive doctrines. This consensus comprises Rawls's idea that the individuals must be able to agree on a political conception of justice like JF, not only from self-interest—and thus not only from an interest in general compliance to the rules of the contract—but on the basis of their personal comprehensive doctrines. While these doctrines certainly differ in many regards, the overlapping consensus is considered possible for two reasons.⁴⁵ First, the political values of all members of a society are very important, as they determine the basic structure of social life. Second, Rawls insists that reasonable comprehensive doctrines have always left room for interpretation, as the development both of Christian and Islamic thought show. Theologians and philosophers have continuously been able to demonstrate how religious and philosophical values are compatible with political values. This is possible, according to Rawls, because a *political* conception does not say anything about the truth of comprehensive doctrines. These doctrines must only accept the existence of other reasonable comprehensive doctrines.

For the problem at hand, it is relevant that an overlapping consensus differs from a *modus vivendi* in the following three respects:

(1) The overlapping consensus does not depend on changing circumstances and is therefore more stable.⁴⁶

(2) The overlapping consensus comprises conceptions of the person and the society and also principles of justice. It is thus much “deeper” than a *modus vivendi*, as it extends into the moral domain. Moral reasons, which are developed on the basis of reasonable comprehensive doctrines, determine the acceptance of the conception of justice.⁴⁷

(3) Rawls defines the overlapping consensus not as a point of convergence, where conflicting interests converge and where people agree “on accepting certain authorities, or on complying with certain institutional arrangements.”⁴⁸ Mere adherence to the rules is thus not enough for the citizens. Rather, what Rawls demands is adherence for the right reasons. According to him, this demand for the use of reason is a functional one, as it is in the interest of social stability.⁴⁹ If, as I will argue, stability could be achieved without an overlapping consensus, then Rawls's argument would be undermined.

For Rawls, the *modus vivendi* will thus not yield sufficient stability. We might be inclined to take this for granted, as the *modus vivendi* seems indeed rather weak. However, Rawls goes on to argue against an intermediate concept that lies between the *modus vivendi* and the overlapping consen-

45. *See id.* at 133–72.

46. *See id.* at 148.

47. *Id.*

48. RAWLS, POLITICAL LIBERALISM, *supra* note 22, at 147.

49. *Id.*

sus—the constitutional consensus. The latter might evolve from the *modus vivendi* and implies agreement on at least part of the values of political liberalism, in particular about the political procedures and some basic rights and liberties.⁵⁰ As a result, the constitutional consensus should give a society at least a little more stability than the *modus vivendi*, even if the sense of justice is not yet in effect.

But Rawls regards the constitutional consensus as still too narrow. It should only be a first basis for a discussion between different political groups with political doctrines of their own.⁵¹ In order to dispute peacefully in public, the groups have to develop political conceptions of justice. They have to learn how to argue and how to convince other groups. This would lead to a more differentiated discussion of questions of justice, in the course of which an overlapping consensus might arise, by agreeing, among other things, on basic rights and liberties.⁵²

Rawls concedes, however, that the “different social and economic interests” might impede this development.⁵³ Under the following two conditions, the conflicts might turn out to be insurmountable: if the different conceptions of justice are “supported by and encourage deeply conflicting political and economic interests,”⁵⁴ and if the conflicts of interests cannot be overcome by a constitution, then “a full overlapping consensus cannot, it seems, be achieved.”⁵⁵ This seems to imply that such conflicts can only be balanced by constitutional rules, not by any moral surpluses. But if that is the case, where is the higher dignity of the overlapping consensus vis-à-vis the constitutional consensus and the *modus vivendi*, which likewise require both the balancing of interests? Is the overlapping consensus nothing but a long-entrenched constitutional consensus?

It is clear that Rawls has more than that in mind. Overcoming conflicts of interests is apparently a necessary, but not a sufficient, condition for an overlapping consensus. The overlapping consensus offers more stability than the two lower levels, but it can only come into effect if the citizens develop a moral surplus, that is, the sense of justice.⁵⁶ However, there remains a problem when trying to separate the different forms of consensus. Rawls goes on to argue that an overlapping consensus is necessary in a functional way to avoid conflicts.⁵⁷ But remember that certain types of con-

50. *See id.* at 158–59.

51. *Id.* at 164–68.

52. *Id.*

53. *Id.* at 167.

54. RAWLS, POLITICAL LIBERALISM, *supra* note 22, at 168.

55. *Id.*

56. For a critical account of this, see Claudia Mills, *Not a Mere Modus Vivendi: The Bases for Allegiance to the Just State*, in *THE IDEA OF A POLITICAL LIBERALISM: ESSAYS ON RAWLS* 193–203 (Victoria Davion & Clark Wolf eds., 2000).

57. He writes that a “constitutional consensus will prove too narrow,” as a democratic people will have to be “sufficiently unified and cohesive” in order to cover, among others, basic matters

licts can impede the overlapping consensus. So is this a vicious circle? Could it be that the very conflicts that are to be avoided by an overlapping consensus might impede it? To put it differently, could it be that the overlapping consensus does not *actually* solve these conflicts but only *seems* to solve them?

Rawls does not make clear how strong a conflict must be in order to impede an overlapping consensus. Does it have to be a very deep constitutional conflict, like that of abolishing slavery in the United States, or would smaller conflicts (e.g., over the construction of new highways) suffice? Rawls would probably respond that this is not a question for philosophical theory, but only for practical application. Nevertheless, in order to make claims about the stability of different levels of consensus, I think a theory of conflicts would be a desideratum. This theory would have to explain which types of conflicts an overlapping consensus could avoid and which types might impede it.

In sum, Rawls regards the sense of justice as an indispensable element in the development of an overlapping consensus that, in turn, could generate sufficient social stability. There is, however, a certain tension in his argument—between the claim that the sense of justice must not be based on interests, on the one hand, and the claim that the overlapping consensus fulfils a social function in avoiding conflicts (which should be in the interests of all), on the other hand.

IV. GAUTHIER: DISPOSITIONS

In his 1986 book *Morals by Agreement*, D. Gauthier tried to develop an account of morals as based purely on self-interest, thus starting with assumptions fundamentally different from those of the previous two approaches. Gauthier starts by stating that moral problems arise due to PD situations.⁵⁸ In a PD situation, the participants recognize that the result of their (rational) actions is quite unwanted by all parties. The solution proposed by Gauthier is that all actors should commit themselves to a certain principle of justice—the principle of “minimax relative concession” (MRC).⁵⁹ In MRC, each individual accepts a rule that minimizes her highest possible relative concessions to others.⁶⁰ Gauthier argues that MRC is to be adopted in a two-step procedure. First, the actors choose MRC from self-interest. Afterwards, however, MRC is to constrain the actors’ self-interest.⁶¹

of justice. Otherwise, “conflict will arise about these.” RAWLS, *POLITICAL LIBERALISM*, *supra* note 22, at 166.

58. See DAVID GAUTHIER, *MORALS BY AGREEMENT* 12 (1986).

59. *Id.* at 157.

60. See David Gauthier, *Political Contractarianism*, 5 J. POL. PHIL. 132, 135 (1997).

61. For the following, see LUETGE, *supra* note 2.

Gauthier tries to show that MRC would be chosen by rational, utility-maximizing actors for the sake of greater benefits in the long run. However, he recognizes that this is not enough and that the problem of compliance to MRC is not solved easily by pointing to its rationality. Gauthier reminds us of “Hobbes’s Foole.” In the *Leviathan*,⁶² the Foole is introduced to show that an actor’s acceptance of a contract is not a sufficient condition to be motivated to comply with the contract. The Foole sees the remaining PD and argues that it would be best for him if all others complied, but he himself could break the contract.

Now, it is important to note that Gauthier does not want to enforce compliance by sanctions and incentives. This would be “a political, not a moral, solution,”⁶³ which would neglect a central difference between morals and interests—“Were duty no more than interest, morals would be superfluous.”⁶⁴ Morals would be more efficient in solving interaction problems because, according to Gauthier, the cost of supervision and enforcement of norms would be reduced to zero.⁶⁵ A voluntary keeping of moral agreements would make at least some institutions unnecessary. Gauthier is thus not against political solutions *per se*, but he hopes to substitute them by “cheaper” means.

Gauthier’s alternative proposal is that all actors should internalize MRC by adopting a disposition to constrain their actions. He distinguishes between two types of actors: straightforward maximizers (SMs) and constrained maximizers (CMs).⁶⁶ While both types maximize their utility, CMs do so under condition of other actors’ utility. CMs adopt a disposition to cooperate, which Gauthier also calls the “idea of mutual benefit.”⁶⁷ A CM complies with mutually agreed norms if she thinks that her expected utility would be positive in the case of general compliance, thereby tolerating at least some degree of free-riding. She will *not* punish defection by defecting herself. An SM, by contrast, will always try to directly maximize his utility without being inhibited by any internalized constraints.

Gauthier shows that in some evolutionary settings, CMs have an advantage over SMs. If CMs can expect to frequently meet other CMs, they can stabilize cooperation within their group and thereby realize large gains from group-internal cooperation. If this leaves the CMs better off than the SMs, the disposition to cooperate could well spread fast within a society.⁶⁸ Gauthier admits, however, that if all other actors within a society are SMs, then the remaining CMs must also behave like SMs.⁶⁹ In this situation,

62. THOMAS HOBBS, *LEVIATHAN* 74–75 (Ernest Rhys ed., J.M. Dent & Sons Ltd. 1914).

63. GAUTHIER, *MORALS BY AGREEMENT*, *supra* note 58, at 163.

64. *Id.* at 1.

65. *Id.* at 164–66.

66. *Id.* at 15–16.

67. *Id.* at 157.

68. *See id.* at 233–67.

69. GAUTHIER, *MORALS BY AGREEMENT*, *supra* note 58, at 181.

morality would have no chance. Moreover, Gauthier states explicitly that a CM is not just a very sophisticated SM that cooperates because she expects greater benefits in the long run.⁷⁰ Rather, she cooperates even if she does not expect positive retribution.⁷¹ She cooperates because her disposition makes her a moral actor. Here, it becomes clear that, contrary to his original intention, Gauthier disconnects morals from advantages and benefits.

CMs and SMs are conceptualized as having different natures. Specifically, CMs are fundamentally different from the *homo oeconomicus*, which implies a “radically contractarian view of human relationships.”⁷² This, according to Gauthier, would commit us to treating other people only as means for (even if mutual) benefits. The *homo oeconomicus* would only have “asocial motivations”⁷³ in that he would not be motivated by the intrinsic value of human relations. This makes him not only asocial, but also irrational, because Gauthier thinks it rational to comply with intrinsic values.

A question that remains is how the internalization of dispositions might take place in practice, with the explicit purpose of maintaining a stable society.⁷⁴ The problem of social stability is not, as I have already mentioned, to be solved by institutions and sanctions, because these are regarded as “unproductive transfer[s].”⁷⁵ In the long run, at least, the individuals would not accept that resources be wasted on unproductive transfers, which could be used in more efficient ways if all complied with MRC. Eventually, Gauthier thinks, all actors must understand that it would be best to stick to MRC.⁷⁶

However, PD situations remain a problem. Even for Gauthier, the central element “in a contractarian theory is not the introduction of the idea of morality, but the step from hypothetical agreement to actual moral constraint.”⁷⁷ What can help actors out of PD situations that persist even after people internalize the necessary dispositions? How can dispositions be an actual moral constraint if some individuals pretend to be CMs but do not actually cooperate with others? If some such covert SMs must be reckoned with, it is quite doubtful whether the CMs can uphold their dispositions. While Gauthier admits that a CM must act like an SM in a population consisting entirely of SMs, he seems to be confident that a sufficiently large number of CMs is enough to make cooperation among them fruitful.⁷⁸ But how many are necessary? While Gauthier cannot be expected to give a fig-

70. *Id.* at 169–70.

71. *See id.*

72. *Id.* at 319 (italics in original).

73. *Id.*

74. *See id.* at 179.

75. GAUTHIER, MORALS BY AGREEMENT, *supra* note 58, at 197.

76. *See id.* at 190–232.

77. *Id.* at 9.

78. *See id.* at 182–84.

ure, he does not even mention the relevant central idea of the PD, whereby one (even one *potential*) defector is enough to destabilize a social arrangement or a moral norm. So if Gauthier took PD situations seriously, he would be forced to abandon this theory and thereby his moral surplus.

However, in the last chapter of *Morals by Agreement*, Gauthier seems to introduce an additional concept: education. This is a rather casual remark, but I think of great importance to his approach:⁷⁹ “[A]n essentially just society must be strengthened through the development of the affections and interests of the young.”⁸⁰ This seems to be the moral, not political, solution that Gauthier has in mind. But it seems to me that it is not an alternative to sanctioning norms. After all, education cannot go without sanctions. Or does Gauthier rely on some concept of anti-authoritarian education? In that case, the PD would remain—how can norms (or dispositions) be enforced if there is no sanctioning mechanism to bring SMs to cooperate?

It becomes clear that Gauthier recognizes the problem of implementation but ultimately underestimates it.⁸¹ The introduction of dispositions as a moral surplus can only be a makeshift, not a systematic, solution.⁸²

V. BINMORE: EMPATHETIC PREFERENCES

In his seminal “Game Theory and the Social Contract,” Ken Binmore has proposed a contractarian approach he calls “naturalistic,” as it relies heavily both on game-theoretic and on sociobiological concepts. A central demand of Binmore’s naturalism throughout his books is “No commitments!” I will discuss this first and then elaborate on Binmore’s concept of “empathetic preferences,” which are supposed to be necessary for social stability in modern societies. However, as I will try to show, these empathetic preferences are fundamentally different from the moral surpluses discussed so far.

Binmore emphasizes that a naturalistic approach must abandon all authorities legitimated by metaphysics. According to him, the vast majority of contemporary approaches in political philosophy—among them, Rawls, Harsanyi, Gauthier, and Nozick—give metaphysical justifications for rules

79. See also Gauthier, *Political Contractarianism*, *supra* note 60, at 148 (stressing the importance of education).

80. GAUTHIER, *MORALS BY AGREEMENT*, *supra* note 58, at 351.

81. To some extent, this has been recognized by Buchanan and Harman, who doubt whether rational actors would accept a distribution as a result of a hypothetical contract. See JAMES M. BUCHANAN, *The Gauthier Enterprise*, in *THE ECONOMICS AND THE ETHICS OF CONSTITUTIONAL ORDER* 195, 195 (1991); Gilbert Harman, *Rationality in Agreement*, 5 *SOC. PHIL. & POL’Y* 1, 1 (1988). However, in my view, it is not the hypothetical character of the contract that poses a major problem, but rather the kind of social structures in effect (PD situations).

82. According to Binmore, Gauthier invents a non-existent enforcement mechanism. BINMORE, *GAME THEORY VOL. 1*, *supra* note 16, at 26–27.

and institutions.⁸³ The common idea of these approaches is that actors can make commitments. Binmore defines a commitment as an “action in the present that binds the person who makes it irrevocably in the future.”⁸⁴ A commitment is thus a “binding unilateral promise,”⁸⁵ (i.e., a promise that ultimately cannot be revoked and therefore—this is my interpretation—is not and does not have to be enforced by sanctions but can obviously be secured in some alternative way). A commitment is not equivalent to a rule, which an actor observes because she expects it to be beneficial to her (at least in the long run).

The major problem with assuming the possibility of commitments is the construction of a plausible mechanism of enforcement.⁸⁶ It is difficult to commit oneself and convince others that one has committed oneself. One possible solution is to provide (financial) “hostages.” For example, a company that has made a commitment to environmental protection may sign a contract in which they commit themselves to paying a fixed amount of money if the commitment is broken. And there are other more subtle mechanisms of securing commitments via reputation mechanisms. These are commitments that are enforced via sanctions, however, and not the kind of commitments that political philosophers like those mentioned above have in mind.

Binmore proposes a different concept: empathetic preferences.⁸⁷ The intuitive idea behind this is that even a *homo oeconomicus* actor can adapt his actions better to that of other actors if he can predict their behavior. Moreover, Binmore distinguishes sharply between a sympathetic and an empathetic preference as follows: Actor A reveals a *sympathetic preference* if it can be deduced from his behavior that he puts himself into actor B’s position *and* adopts B’s preferences.⁸⁸ By contrast, actor A reveals an *empathetic preference* if it can be deduced from his behavior that he puts himself into actor B’s position *without* taking on B’s preferences. In this case, A sticks to his own preferences. He can still compare his preferences to B’s and evaluate or criticize the latter. To cite Binmore’s comparative preferences example, “I would rather be Eve eating an apple than Adam wearing a fig leaf.”⁸⁹

83. *Id.* at 161.

84. *Id.*

85. *Id.*

86. *See id.* at 162.

87. This goes back to Harsanyi’s “extended preferences.” *See* JOHN C. HARSANYI, RATIONAL BEHAVIOR AND BARGAINING EQUILIBRIUM IN GAMES AND SOCIAL SITUATIONS 65 (1977). By employing the concept of ‘empathetic preferences’, Binmore wants to highlight the difference between the classic concept of sympathy in D. Hume’s works and his modern one. *See* BINMORE, GAME THEORY VOL. 1, *supra* note 16, at 28, 58–60, 285–96; BINMORE, GAME THEORY VOL. 2, *supra* note 16, at 212.

88. *See* BINMORE, GAME THEORY VOL. 1, *supra* note 16, at 286.

89. *Id.* at 290.

Binmore argues that the ability to empathize,⁹⁰ not the ability to sympathize,⁹¹ with others is what makes a human being. He speculates about the evolutionary history of this ability, which may have been advantageous for coordinating behavior in hunter-gatherer societies.⁹² And he distinguishes between three time horizons in which empathetic preferences play a role: short, medium, and long run.⁹³

In the short run, the actor's personal preferences, as well as her empathetic preferences, are fixed.⁹⁴ That is, the actor empathizes in exactly the way her empathetic preferences prescribe and deliberates "morally" in this way. Here, moral norms are conventions that work as short cuts for long economic calculations. While morals do not play a role on the level of the social contract framework, they do influence the individuals' actions within this framework. Morals are functional in the short run.⁹⁵

In the long run, all preferences, personal and empathetic, are subject to change, as the actor adapts to new situations and new rules.⁹⁶ Here, new social contracts are negotiated and existing ones modified. The personal and empathetic preferences adapt to these new situations. It is interesting to see that in the long run, all moral content erodes out of the preferences. Over longer periods of time, the actors arrive—via "moral" empathy—at the same result as if they had been bargaining straightaway all the time. Binmore makes it clear that in the long run, morals serve long-run interests, and more importantly, no moral norms can remain stable that are systematically opposed to incentives.⁹⁷

Finally, in the medium run, the personal preferences remain fixed while the empathetic ones may change.⁹⁸ According to Binmore, evolution will bring the latter into "empathy equilibrium,"⁹⁹ in which all actors have equal empathetic preferences. It is already in this situation, as in the long run, that all moral content erodes from the social contract framework. The

90. I prefer to use the term 'ability to empathize,' as it is a more general concept than the empathetic preferences. The latter are preferences that a particular actor reveals in a concrete case with regard to one or several other actors. These may change from one actor to another and from case to case.

91. But according to Elster, 'love and duty,' i.e., sympathy in Binmore's terms, are the "cement of society." JON ELSTER, *THE CEMENT OF SOCIETY* 287 (1989). Binmore responds that modern societies do not need cement, rather they are like a dry-stone wall in which each stone is kept in place only by the other stones, i.e., by reciprocity. BINMORE, *GAME THEORY VOL. 1*, *supra* note 16, at 24. To maintain reciprocity, however, greed and fear suffice.

92. See BINMORE, *GAME THEORY VOL. 1*, *supra* note 16, at 57, 288–90.

93. *Id.* at 86.

94. *Id.*

95. See BINMORE, *GAME THEORY VOL. 2*, *supra* note 16, at 449–50.

96. See BINMORE, *GAME THEORY VOL. 1*, *supra* note 16, at 90.

97. See *id.* at 89–90.

98. *Id.* at 87–88.

99. *Id.* at 65; see also *id.* at 61–67, 86–90, 290–92.

actors end up with a result identical to a Nash bargaining equilibrium.¹⁰⁰ However, this is at variance with the actors' own impression. Due to the semantics employed, they still think that they are guided by moral deliberation in the traditional sense. Binmore consequently regards the ability to empathize as—at least partially—genetically “hard-wired.”¹⁰¹

The difference between the long and medium run is that in the medium run, evolution has not had enough time to shape personal preferences and adapt them to new environments and situations. By distinguishing between the three time horizons, Binmore defines the role of empathetic preferences in society—they are used for coordination or, more precisely, for reforming existing social contracts and consenting to new ones. They are used as a heuristic tool for finding directions in which new social contracts may develop. In this sense, Binmore regards the empathetic preferences as an important part of morality.¹⁰²

The question now is whether empathetic preferences can be regarded as a moral surplus in the sense of Habermas's rational motivation or Rawls's sense of justice. I think not, for the following reasons. Upon closer inspection, it becomes clear that empathetic preferences have fewer consequences for their bearers than the other moral surpluses discussed here. In Rawls's and Habermas's works, one gets the impression that both authors already have in mind a rather precise idea how the citizens should act—or at least which rules they should adopt. In addition, both works show clear traces—albeit to different degrees—of an opposition to economic approaches or, more generally, to those that rely on self-interest for their implementation.¹⁰³

This impression does not arise when reading Binmore. First, the empathetic preferences exhibit a peculiar quality—assuming their existence does not preclude anything for the detailed design of rules and institutions.¹⁰⁴ If A can put herself in B's position, she will in some way try to assess B. On the one hand, if A regards B as rather unreliable or as only interested in short-run gains, she will anticipate B's defection, adapt her

100. See *id.* at 88. For the Nash equilibrium, see John F. Nash, Jr., *Equilibrium Points in n-Person Games*, 36 PROC. NAT'L ACAD. OF SCI. 48, 49 (1950), and John F. Nash, Jr., *The Bargaining Problem*, 18 ECONOMETRICA 155 (1950).

101. BINMORE, GAME THEORY VOL. I, *supra* note 16, at 133; BINMORE, GAME THEORY VOL. 2, *supra* note 16, at 182. He does not rely entirely on genetic concepts, however, but uses the concept of the “meme.” RICHARD DAWKINS, THE SELFISH GENE 189 (30th Anniversary ed. 2006); see BINMORE, GAME THEORY VOL. 1, *supra* note 16, at 65–66.

102. BINMORE, GAME THEORY VOL. 1, *supra* note 16, at 241.

103. Compare Rawls's strict separation between the *modus vivendi*, the constitutional consensus, and the overlapping consensus (see *supra* Part III).

104. However, this can also be seen as a shortcoming of Binmore's work—he does rely exclusively on evolutionary game theory, as Dore also emphasizes. Mohammed Dore, *On Playing Fair: Professor Binmore on Game Theory and the Social Contract*, 43 THEORY & DECISION 219, 236 (1997). While this is certainly relevant to modern societies, we cannot however go without other branches of economics and other social sciences for the detailed design of institutions.

own behavior, and tend to “counter-defect pre-emptively” herself.¹⁰⁵ If, on the other hand, A regards B as reliable *and* if there are no contrary incentives for A to defect, A will tend to cooperate. But this is an open-ended process, the result of which cannot be precluded by the philosopher.

Second, and more important in my mind, the ability to emphasize cannot, unlike the moral surpluses reconstructed here, *be exploited* by other actors. Consider this: If A constrains her behavior in a PD situation (e.g., by subscribing to a rational motivation or a sense of justice), she risks being exploited by B. This can only be avoided if B constrains his behavior in the same way as A (i.e., by way of sanctions or the like). But if A can just empathize with B in a PD situation (i.e., rely on empathetic preferences), she does not necessarily risk exploitation, especially not in a situation where B acts in ways different from A. For example, suppose that one of the two prisoners in the classic PD situation (X) is motivated by a sense of justice. If the other (Y) knows this but is not motivated in this way himself, he can exploit X without any problem by confessing (i.e., defecting). But if X “only” has empathetic preferences, this does not necessarily lead to exploitation. It would only mean that X might anticipate the reaction of Y to his own “moral” behavior. X could, for example, use this knowledge to try to turn the tables and exploit Y. In any case, Y cannot gain any unilateral advantage from knowing that X has adopted empathetic preferences, as Y would have to count on the fact that X would anticipate this—by empathizing with Y.

Therefore, I suggest not putting the ability to empathize in the same basket as the moral surpluses discussed above. It does not preclude any particular action, and it cannot be exploited.

VI. ORDER ETHICS AS AN ALTERNATIVE VIEW: STABILITY FROM ADVANTAGES AND INCENTIVES

The idea behind the approach that I would like to sketch here is to develop an account of norms as functional for social stability while weakening the assumptions on the actors’ part (i.e., minimizing the necessary moral surplus). This alternative view is a conception of ethics that proceeds systematically not from the problem of justification but from the problem of implementation.¹⁰⁶

105. See KARL HOMANN, *VORTEILE UND ANREIZE: ZUR GRUNDLEGUNG EINER ETHIK DER ZUKUNFT* 98 (Christoph Luetge ed., 2002); KARL HOMANN & CHRISTOPH LUETGE, *EINFÜHRUNG IN DIE WIRTSCHAFTSETHIK* 35 (3d ed. 2013).

106. For the order ethics approach, see LUETGE, *supra* note 2; ORDER ETHICS: AN ETHICAL FRAMEWORK FOR THE SOCIAL MARKET ECONOMY (Christoph Luetge & Nikil Mukerji eds., forthcoming); HOMANN, *supra* note 105; HOMANN & LUETGE, *supra* note 105; Christoph Luetge, *Economic Ethics, Business Ethics and the Idea of Mutual Advantages*, 14 BUS. ETHICS: EUROPEAN REV. 108, 110–11 (2005). The corresponding idea in Binmore’s approach is the requirement to *first* regard only the relevant equilibria and *then* pick one of them as the desirable.

Some philosophers, both in the analytic as well as in the continental tradition, believe that the justification of norms has always been the key problem of ethics. However, an informed view of economic and social history¹⁰⁷ tells us that the questions of norm implementation *and* justification have generally been posed together, but this was not made explicit in former centuries. The implementation of norms that had already been justified was not regarded as particularly difficult for two reasons. First, the modern pluralism of values and life styles had not yet developed, and second, social relations were not yet as anonymous as in modern times. In particular, commonly accepted norms could be enforced much more easily through face-to-face sanctions.¹⁰⁸

Since the beginning of modern times, however, this situation has changed dramatically. As N. Luhmann has described, modern societies consist of functionally differentiated sub-systems.¹⁰⁹ The actors in modern societies must act in social sub-systems under completely different governance mechanisms. With regard to ethics, this often leads to laments about loss of values. The question how norms can be enforced therefore becomes much more pressing under modern conditions and must be put at the beginning of a conception of ethics for modern societies. I would like to call such a conception an *order ethics*, or in a different regard, an *ethics of advantages and incentives*. We must count on the fact that all actors look for their advantage (whereby the term “advantage,” of course, covers not only material advantages), and we have to set the right incentives in order to deal with this quest for advantages.

Order ethics can be set in a contractarian framework and derived from a thought experiment. For this argument I refer to B. Ackerman, who has renewed the classic social contract argument with the aid of a science-fiction thought experiment:¹¹⁰ A group of spacemen lands on a newly discovered planet and tries to build a society there. The spacemen have to deliberate about the rules of this society in much the same way as the rational actors in the social contract setting of an original position. Ackerman, however, explicitly uses the science-fiction scenario in a very peculiar way, namely to get rid of the question how rules that have been agreed on can be effectively enforced.¹¹¹ He assumes that in his fictitious scenario, violations of rules can be sanctioned automatically by means of a superior technology (laser cannons).

107. For such a view on Ancient Greece, see generally CHRISTIAN MEIER, *ATHENS: A PORTRAIT OF THE CITY IN ITS GOLDEN AGE* (Robert Kimber & Rita Kimber trans., Metropolitan Books 1998) (giving a historical overview of ancient Athens).

108. See LUETGE, *supra* note 2.

109. See, e.g., NIKLAS LUHMANN, *DIE GESELLSCHAFT DER GESELLSCHAFT* (1997).

110. See BRUCE A. ACKERMAN, *SOCIAL JUSTICE IN THE LIBERAL STATE* (1980).

111. *Id.* at 34.

I do not think this approach leads us very far. Therefore, I have changed Ackerman's setting, because the perfect enforcement of rules via technology simply is not available, even in a futuristic world. If one drops the assumption that there are laser cannons guaranteeing compliance, Ackerman's thought experiment can be sensibly taken further. Also, I add the idea that the people aboard the ship already have very different values and backgrounds. The result is a situation in which rules have to be designed for a new situation in a society with great pluralism.

From this revised thought experiment, the main thesis of an order ethics can be developed in four steps. First, the problem must be clear for which an order ethics might be a solution. This problem is the *problem of social order*.¹¹² Second, the problem cannot be solved by way of an *individual ethics*. An individual ethics (ideally) assumes that morally problematic states are caused by actors' immoral motives or preferences. Consequently, this position calls for a change of motives, or a change of consciousness. The main mechanism of governance is (moral) appeals, and may be supported by education.

However, in those structures typical of modern societies (i.e., PD situations),¹¹³ an ethics remains fruitless that addresses primarily the individual. So in a third step, an order ethics assumes that morally problematic states are not caused by immoral preferences or motives but by specific structures of interaction. Therefore, moral claims should aim at revising the conditions that apply to all actors (i.e., the rules of the game). The main governance mechanism is the design of *incentive* structures. Moral norms cannot be brought into opposition to the logic of advantages and incentives.

Fourth and finally, some refinements have to be made to the social contract framework. I limit myself to the following two: Contrary to what some critics of this approach believe, rational actors can invest in the future. They can invest in better rules, but they can also invest in moral behavior. There is still room for moral behavior in a social-contract-based order ethics approach, but this behavior can be explained with the help of the concept of incomplete or open contracts. As many contracts in modern societies are systematically incomplete,¹¹⁴ the actors must fill in the gaps in contracts with their own "moral" behavior—in their own interest.

The arguments in the preceding sections make it clear that it is not systematically fruitful to base normativity on anthropological capabilities or characteristics—the moral surpluses. But if these moral surpluses are not

112. See generally FRIEDRICH A. HAYEK, *LAW LEGISLATION AND LIBERTY VOLUME 1: RULES AND ORDER* 35–54 (1973).

113. As in Gauthier and Binmore, PD situations occupy a pivotal place in this approach.

114. See Oliver Hart, *Incomplete Contracts*, in *THE NEW PALSgrave: A DICTIONARY OF ECONOMICS* (John Eatwell, Murray Milgate, & Peter Newman eds., 1987); Oliver Hart & Bengt Holmström, *The Theory of Contracts*, in *ADVANCES IN ECONOMIC THEORY – FIFTH WORLD CONGRESS* 71 (Truman F. Bewley ed., 1987); Luetge, *supra* note 106, at 113.

sustainable, what are the alternatives? The alternative social governance mechanism that an order ethics would argue for is not based on anthropological findings, characteristics or surpluses, but on situations. A modern society that wants to profit from deep specialization and competition has to switch to a system of governance by rules. With the use of pre-modern governance mechanisms, the benefits of modern societies cannot be appropriated.

Which rules and which governance mechanisms are necessary for an interaction depends only on the situational conditions. For example, there might be situations in which informal governance by moral norms still works, and the partners can count on at least approximately equal normative backgrounds. In such situations, governance by moral norms may still be an option. However, these cases are not very frequent in modern societies under conditions of globalization. The number of interactions between individuals with vastly different cultural, social, and normative backgrounds is continually and rapidly growing. These individuals cannot rely on common moral surpluses; they can only rely on common mutually accepted rules—or they might devise new rules adapted to their situation.

The question is whether even for these cases of rule governance some—maybe weak and not anthropological, but rather situational—capability is necessary. Of the concepts discussed here, only the ability to empathize would be a candidate. Notwithstanding this, I argue that if we abandon any moral surpluses, there are three minimal—and very general—assumptions that must be given to guarantee functional governance by rules and social stability: sociality, ability to communicate, and ability to invest.

The first two are rather trivial. That is, there must be some social group and some mode of communication in effect within this group. But the third is more interesting and important. Individuals that always maximize utility in the *short* run cannot form a stable society. This is nothing spectacular and is not peculiar to social stability or morals. Every company must be able to invest in the future. In fact, any form of action and cooperation requires thinking and planning, which in turn requires investing (i.e., saving some resources now for greater benefits in the longer run).

Together, these three assumptions might be taken as a minimal basis for modern globalized societies that cannot rely on moral surpluses. The actors must only be able to communicate and invest. An order ethics can then go on to only require these actors to one, comply with the rules, and two, engage from their own interest in the further development of these rules in mutually beneficial ways. Such further development can only come into effect if the individuals affected agree to it. In light of PD situations, it is systematically not enforceable against the wishes of these individuals.

However, there is a role for what I would like to call *heuristics*. These heuristics are values and ideas from philosophical, religious, scientific, literary, artistic, or other traditions. They can point the actors in new direc-

tions where new gains of cooperation may be found. But they are conceived here in a very different way than the moral surpluses. Most importantly, they are not conceptualized against the logic of advantages and incentives. They do not erode in PD situations. By contrast, the moral surpluses that Habermas, Rawls, and Gauthier argue for are all conceptualized in opposition to interests, advantages, or incentives.

To sum up, an order ethics starts in much the same way as Gauthier and Binmore suggest. All three approaches regard moral norms in a functional way—as a possible means for governing societies. All three stress the importance of interactions, and all three see no fundamental contradiction between morals and self-interest. However, the order ethics approach aims at solving the problem of implementation primarily by means of rules or institutions, not (unlike Gauthier) by dispositions, and not (unlike Binmore) solely by the ability to empathize. Maybe Binmore would agree with my approach, as he is not at all opposed to employing institutions for implementing norms. However, they are not what he focuses on. I would therefore think that order ethics and Binmore's evolutionary approach are two complementary conceptions.

The theoretical advantage of both the ability to empathize and the ability to invest lies in the fact that they do not require the individuals to act against the logic of advantages and incentives, especially in PD situations. A practical application might be that a “morally driven” semantics used in the political sphere, which constructs sharp contradictions between values and interests, between rational motivation and incentives, between acting from principles of justice and acting “merely” in accordance with them, leads us to dead ends and theoretical blockades. If, however, a revised semantics employs concepts like “investing,” “mutual self-interest,” “mutual gains,” or “win-win-situations,” then even the moral surpluses might be made productive. They might be seen as part of a heuristics which asks for, suggests and pleads for, *investing* in the individuals' own self-interest. Eventually, this could turn out to be the heuristics of not only occidental but also many other traditions of ethics. To show this in detail, however, is a task for the future.