

*David Copp*

## **Comment on Lorenzo Sacconi, Marco Faillo and Stefania Ottone. Contractarian Compliance, Welfarist Justice, and Conformist Utility**

*Abstract:* This comment addresses two issues that arise in Sacconi/Faillo/Ottone's essay. The first is the problem of compliance as it arises in social contract theory. The second is the problem of avoiding an incoherence that arises in the formulation of welfarist principles of distributive justice if these principles are taken to be concerned with the distribution of welfare without restriction. Sacconi, Faillo, and Ottone define an interesting class of principles that govern only the distribution of 'material utility', which they distinguish from 'conformist utility'. Sacconi, Faillo, and Ottone are primarily concerned, however, to argue that there is a need to revise 'the utility maximization model of a rational economic man'. I discuss this claim briefly, in concluding the paper.

### **1. Introduction**

Sacconi, Faillo, and Ottone contend that there is a need to revise "the utility maximization model of a rational economic man" (292). Following suggestions of John Rawls (1971), they argue that people can have a "conformist preference", and that in light of this, a person's "comprehensive utility function" has two aspects, which they term "material utility" and "conformist utility" (292). They discuss two experiments that support their claim about the conformist preference. I believe that matters are more complicated than this. But I will approach the question indirectly, by discussing two other issues that arise in Sacconi, Faillo, and Ottone's article. I begin with the problem of compliance as it arises in social contract theory. This is the problem of showing that it would be rational for persons to comply with norms that they would be rational to agree in advance to comply with. The second is the problem of avoiding an incoherence that arises if welfarist principles of distributive justice are taken to be concerned with the distribution of welfare without restriction. Sacconi, Faillo, and Ottone define an interesting class of principles that govern only the distribution of material utility. My discussion of this class of principles will lead me back to Sacconi, Faillo, and Ottone's claims about moral psychology.

## 2. The Problem of Compliance

One traditional question in moral philosophy is whether there is any guarantee that a rational person will be moral. The question arises in an acute form in the neo-Hobbesian tradition, which, as I understand it, assumes that rational agents do or try to do what is best for themselves. The problem is that, on an orthodox view, morality sets constraints on the permissible pursuit of one's own good. I am not permitted to lie or cheat or harm or kill people simply in order to bring about some greater good to myself. If this is right, then given the contractarian assumption about the nature of rationality, it follows that rationality and morality can conflict. It may not be morally permissible to do what would be best for oneself. Given the orthodox view of morality and the Hobbesian assumption about rationality, nothing *guarantees* that a rational person will be moral. Indeed, on these two assumptions, one would expect rational agents to be disposed to act immorally whenever doing so promises them a benefit.

The Hobbesian assumption about rationality is widely shared. It is accepted in one form or another by contractarians, neo-Humeans, and some virtue theorists. Call it the maximizing assumption about rationality. It needs to be formulated more carefully. The assumption is not that a rational agent must have the achieving of what would be best for herself as the only thing or the most important thing that she pursues for its own sake. A rational agent might pursue a variety of things, each for its own sake, including knowledge, friendship, power, and pleasure. The idea, however, is that the rational action for a person in a given context is the action the performing of which promises to be best for her. She might not realize that this action promises to be best for her. She might think of it as what she most wants to do, or she might think of it as the thing that will be best for her children. But if she acts rationally, this is the action she will perform. Moreover, on the maximizing assumption, a rational person is so disposed that, in every situation, she has a tendency to do the action that promises to be best for her.

Philosophers have attempted of course to defend morality against the charge that acting morally might well be irrational. One strategy is to argue against the maximizing conception of rationality. A second strategy is to retain the maximizing conception but to argue for an expansive and morally-loaded conception of what would be best for a person. Plato argued in *The Republic* that an agent's good consists in having a well-ordered psychology, and that a person with such a psychology would be morally just. Some virtue theorists contend that acting morally is always best for the agent because virtue is its own reward, or because the virtuous are always as happy as they could be. A third strategy is to argue for an unorthodox and dubiously narrow conception of morality on which there can be no conflict between morality and self-interest. This is the approach of the ethical egoist. My purpose in this first part of my paper, however, is to examine the approach that is suggested by Sacconi, Faillo, and Ottone.

Sacconi, Faillo, and Ottone are working in the social contract tradition as it has been formalized and developed in the formal theory of rational choice, that

is, in decision theory and game theory. The underlying ideas of this tradition are two-fold. First, no norm or social institution is legitimate or justified unless those who will be regulated by it would rationally consent to be regulated by it, and would rationally comply with its requirements. As Sacconi, Faillo, and Ottone say, “the social contract approach maintains that norms and institutions must be based on the consensus and voluntary compliance of those regulated by the norm itself” (273). Second, the rational action for an agent in any situation is the action that promises to contribute most to the satisfaction of her desires. This is the neo-Humean reading of the maximizing conception of rationality.

These underlying ideas take specific and more precise forms in the formal theory of rational choice. To make precise the neo-Humean view of rational choice, a utility function is defined in mathematical terms, such that an agent’s utility for a given state of the world is a function of the agent’s preference satisfaction in that state of the world. The rational choice for a person in a situation is then said to be the action, of all those available to the agent in that situation, such that the agent’s performing it maximizes the agent’s expected utility. And to make precise the justification requirement for norms, it is specified that a norm or social institution is justified only if two conditions are satisfied. First, any agent who is unfailingly rational in the stipulated sense would agree in advance (*ex ante*) to comply with that norm or institution, at least in an appropriate specified context. Second, assuming the first condition is satisfied, any agent who is unfailingly rational in the stipulated sense would comply with it (*ex post*) even in the absence of any mechanism of enforcement. These are two separate conditions. Satisfaction of the first does not entail satisfaction of the second.

From the standpoint of the formal social contract tradition, the compliance problem is the problem of ensuring the rationality of *ex post* compliance. On the social contract approach, solving the compliance problem is critical to showing that a system of norms is justified. From this perspective, unless we can show that it is rational to comply with the requirements of morality, morality is not justified. Unless we can show that there are requirements of justice with which it is rational to comply, requirements of justice are unjustified.

We can see by inspection, however, that the compliance problem is insoluble given the way it is defined in the social contract tradition. To see this we can set aside the question of whether rational agents would agree *ex ante* to comply with a norm. For the only question is whether the norm might require agents in certain circumstances to do things that would not maximize their expected utility. If it would *not* require such things, then the compliance problem is trivially solved, but the norm is uninteresting. It amounts merely to a requirement to act rationally. But if the norm might require agents in certain circumstances to do things that would not maximize their expected utility, there are possible circumstances in which a fully rational agent would not comply with it. The norm might require truthfulness. But we can easily imagine circumstances in which telling the truth would not maximize an agent’s expected utility. The norm might require sharing with others. But we can easily imagine circumstances in which sharing with others would not maximize an agent’s expected utility

Sacconi, Faillo, and Ottone say,

“If the game within which players must decide to comply with the social contract is a [Prisoner’s Dilemma], as seems unavoidable if the ‘state of nature’ wherein they are embedded is a PD, then the social contract will obviously not be complied with, since non-compliance (naturally corresponding to mutual defection) is the only equilibrium point.” (275, also 278)

It would take me too far afield to explain exactly what they mean here. Suffice it to say that they are agreeing basically with what I have just said, that the compliance problem is insoluble as defined. They add,

“Some additional features of the game must be added to the representation of the state of nature, which by itself necessitates a sub-optimal solution of the noncooperative interaction played by selfish and rational players.” (276)

What they mean is that since the compliance problem is insoluble as defined, we must change how we are conceiving of it, by changing something in our description of the circumstances (the ‘state of nature’) in which we imagine agents to be deciding whether to agree *ex ante* to comply with a norm and to be deciding as well whether actually to comply with it *ex post*.

This brings me to Sacconi, Faillo, and Ottone’s interesting proposal. They have noticed John Rawls’s hypothesis that people are moved by a sense of reciprocity and fairness and that, given this, under favorable circumstances, people can be expected to develop an attachment to principles of justice and a motivation to comply with them (Sacconi et al. 289, quoting Rawls 1971, 491, 494). This means, they suggest, that people have “a deep psychological tendency to answer in kind”, such that, in favorable circumstances, if we see that other people are conforming with principles of fairness, and if we take their conformity to express “an intention to be beneficial to us in an unconditional and impartial manner”, we will tend to conform as well and we will develop a sense of justice. Being just can in this way become “an integral part of what we see as our good” (289).

What, then, are the favorable circumstances in which this tendency would motivate a person to comply with a principle? Sacconi, Faillo, and Ottone find in Rawls the suggestion that this tendency is more likely to be activated when two conditions are satisfied. First, it is more likely to be triggered when the person sees or expects that the others with whom she is interacting are conforming to the principle. In this case, the person will be reciprocating conformity. Second, the tendency is more likely to be triggered when the principle in question has been accepted by the person and by the others with whom she is interacting in an “impartial agreement under the ‘veil of ignorance’” (290). That is, roughly, it is more likely to be triggered when the person herself and all people with whom she is interacting have agreed to be bound by the principle in a situation in which none of them knew how the principles would affect themselves personally. Our

tendency to comply with a principle is not triggered merely by others' compliance with some principle or other, no matter how arbitrary it might be. It is triggered only if we take the principle to have been agreed to by all concerned in a fair and impartial situation behind a veil of ignorance.

Given all of this, Sacconi, Faillo, and Ottone suggest a reconceptualization of the compliance problem. Instead of requiring a theorist to show that a desire to comply with a principle can be "the output from [...] the decision theoretic machine", allow rather that "the desire to comply could be an input to the compliance decision". Of course, we would then "need to understand how this desire may be engendered and how it is connected to the social contract". Rawls's suggestions about the development of the sense of justice lead Sacconi, Faillo, and Ottone to look for "causal connections between the decision to comply and the rationality of an *ex ante* agreement" (284). These connections are cemented by the fact that, as they think, our desire to comply with a principle is triggered only if we take the principle to have been agreed to by all concerned in a fair and impartial situation behind a veil of ignorance.

It is of course no surprise that the compliance problem facing social contract theory is less hopeless if we postulate that people have a desire to act morally or to comply with principles that are widely complied with. But Sacconi, Faillo, and Ottone's proposal is more subtle than this. For in the first place, their proposal links the *ex ante* problem as it is understood in the social contract tradition with the *ex post* problem of compliance. They propose that people have a tendency to comply with principles in interacting with others when (roughly) the principles have been agreed to both by them and by the others in an initial fair situation 'behind a veil of ignorance'. That is, the principles in question must first pass a contractarian test of being such that those putatively bound by them have agreed to be bound by them. And in the second place, their proposal is empirical. It has been tested in experiments that have shown it to have some empirical plausibility.

I should be cautious here. In two experiments with the 'Exclusion Game', Sacconi, Faillo, and Ottone observed a high degree of compliance with distributive rules that had been chosen in advance by the relevant participants. (This game and the design of the experiments are described at 295–302.) In both the experiments, in the circumstance in which the rule had *not* been chosen in advance, participants tended to act in a self-interested manner. In one, 59.3% and in the other, 73.8% of the relevant players acted in a self-interested manner. But in the circumstance in which the rule *had* been chosen in advance, in one experiment 77% of the players chose to conform with the rule even though it was to their material advantage not to conform and in the other 50% chose to conform. Hence, Sacconi, Faillo, and Ottone say, "in both the experiments, for a significant percentage of subjects, agreement on a fairness rule seems to have been a sufficient condition for the emergence of expectations of reciprocal conformity" (304). It is also true of course that in both experiments a significant percentage of subjects failed to conform to a rule that had been chosen in advance, and instead acted to promote their material self-interest.

It should now be clear that Sacconi, Faillo, and Ottone have not shown us a way to solve the compliance problem. As it arises in the social contract tradition, the problem, again, is to show that rational agents would comply *ex post* with norms of justice or morality or, for that matter, with any norms that might require them to do otherwise than to maximize their expected utility. The tradition holds that a norm is justified only if (1) any agent who is unfailingly rational in the relevant sense would agree *ex ante* to comply with that norm, at least in an appropriate specified context and (2) any agent who is unfailingly rational in the relevant sense would comply with it *ex post* even in the absence of any mechanism of enforcement. Sacconi, Faillo, and Ottone have not shown that any norms meet these conditions. What they have shown, instead, is that, roughly, *some* agents will reciprocate expected conformity with a principle that requires them to do otherwise than to serve their *self-interest* when they and the agents with whom they are interacting have themselves *actually agreed* to be bound by the principle. This is an interesting result, but it falls short of solving the compliance problem, as I am sure Sacconi, Faillo, and Ottone understand.

First, the experiments put subjects in situations in which they *actually* have agreed on a principle. The fact that many subjects in such a situation desire to comply with the principle does not speak to whether people would have a desire to comply with a principle that had *not* actually been agreed to, even if it *would be rational* to agree to it. Hence the experiments, and Sacconi, Faillo, and Ottone's hypothesis about the sense of justice, simply do not address the problem of providing a contractarian justification of ordinary moral principles, and principles of justice such as Rawls's, which have not actually been agreed to. Second, in the experiments, many agents *failed* to conform to the principle that had been agreed to. So the experiments do not tend to show that *all* agents who were party to such an agreement on a principle would comply with it *ex post*. Third, as far as the experiments show, it is possible that all the participants acted to maximize their expected utility, including those who complied with the agreed principles *ex post*. Sacconi, Faillo, and Ottone are assuming that people can in some circumstances get extra utility simply by conforming to an agreed principle, for they are assuming that people have a desire to comply reciprocally with principles in interacting with others when the principles have been agreed to both by them and by the others. The experiments do provide some support for this hypothesis. But what this means is that it is possible that no-one in the experiments complied with a norm in circumstances in which it required them to do otherwise than to maximize their expected utility. Hence the experiments cannot help to show a way in which a norm might be justified that requires agents in some circumstances to do things that fail to maximize their expected utility.

### 3. The Problem of Avoiding Incoherence in the Formulation of Social Welfare Functions

The most important suggestion in Sacconi, Faillo, and Ottone's paper, I think, is their suggested bifurcation of agents' utility functions. (In this and the next paragraph, I present in nontechnical terms the account found at 293.) They suggest a formulation that separates an agent's utility into two components, her "material utility" and her "conformist utility". Their innovation is the idea that agents have a "conformist utility" as well as a material utility. As I will explain, an agent's conformist utility answers to the strength of her "conformist preference", her desire to conform her behavior reciprocally with distributive principles that have been agreed to by all concerned. An agent's "material utility" is a function of the degree to which her "material preferences" are satisfied, where her material preferences exclude the "conformist preference". The overall or net utility of a person in a situation is the sum of the person's material utility in the situation and the person's conformist utility in the situation.

According to Sacconi, Faillo, and Ottone's definition, a person has a conformist utility for a state of the world only if some "social welfare function"  $T$  that governs the distribution of material utilities in the relevant population has been accepted unanimously in the relevant population (293). In this case, a person's conformist utility for a situation depends on two chief factors. First, her conformist utility in a situation is larger as, to a larger degree, she and other people in the situation act reciprocally in such a way as to bring about the distribution of material utilities required by  $T$ . And second, given the degree to which she and other people in the situation interact reciprocally to satisfy  $T$ , her conformist utility for the situation is larger as, to a stronger degree, she is motivated by the desire to reciprocate conformity with principles that have been accepted by all concerned. So a person's conformist utility in a situation where some principle  $T$  has been accepted by all concerned is a function of the degree to which  $T$  is fulfilled by reciprocal conformity and the degree to which the person wants to reciprocate conformity.

One nice thing about Sacconi, Faillo, and Ottone's proposed bifurcation of agents' utility functions is that it takes a step toward answering worries about the coherence of welfarist principles of distributive justice that have been raised by Ronald Dworkin (2000, esp. ch. 1). Dworkin notes that the idea of welfare can be interpreted in different ways just as the idea of a person's own good can be interpreted in different ways. But let us confine attention to preference satisfaction theories of welfare of the kind that are accepted in neo-Humean and contractarian theories and in formal theories of rational choice. On a preference-satisfaction theory, a person's welfare or utility is enhanced, other things being equal, by the satisfaction of *any* desire she has, including desires for the good of others and desires for external and distant states of the world. Moral and idealistic desires and immoral and cynical desires are all included, including the egalitarian desire to see welfare distributed equally and the aristocratic desire for an inegalitarian hierarchy of welfare distribution. Now consider unrestricted

welfarist principles of distributive justice, principles that govern the distribution of overall welfare without restriction.

To begin, consider an unrestricted egalitarian welfarist principle. Suppose that I have the egalitarian desire and you have the aristocratic desire. Then as our society becomes more egalitarian, I gain utility due to the increased degree to which my egalitarian desire is satisfied and you lose utility due to the increased degree to which your aristocratic desire is frustrated. It follows that, other things being equal, an unrestricted egalitarian welfarist principle would require compensating you and taxing me in order to correct for the added benefit to me of increasing equality. This is paradoxical. An egalitarian principle should not require taxing egalitarians to compensate inegalitarians who regret the achieving of equality. The problem is not exclusively a problem for welfare egalitarians. Consider an unrestricted prioritarian principle, a principle that requires giving priority to helping the worst-off in society, as measured by persons' unrestricted welfare. Suppose that you have an enormously strong desire for a specific kind of hierarchy and suppose that this desire is frustrated due to the fact that the preferred hierarchy does not exist. You lose utility as a result. In this case, you might be among the worst-off in society as measured by your overall welfare even if you otherwise enjoy much greater welfare than I do, having better housing, a better diet, better health, and so on. In this case, other things being equal, an unrestricted prioritarian principle would require giving priority to helping you over helping me. It might require taxing me in order to help you. This again is paradoxical. A prioritarian principle should not require taxing people who are otherwise worse-off in order to compensate people who are otherwise very well-off and only count as among the worst-off because they have strong unjust preferences that are being frustrated.

Dworkin's strategy for avoiding worries of this kind is to abandon welfarism. But one might instead attempt to avoid the worries by abandoning the idea that distributive justice requires conformity to any *unrestricted* welfarist principle. That is, one might amend the egalitarian view and other welfarist principles so that they do not govern the distribution of *overall* welfare, without restriction. Dworkin contends that any amendment of this kind will be open to objections of similar kinds. But an obvious approach would be to distinguish between 'distributional preferences' and 'ordinary preferences', where a person's distributional preferences concern the distribution of goods in the society and where all other preferences are counted as ordinary. We could then define a person's 'ordinary welfare' or utility as a function of the degree to which her ordinary preferences are satisfied. Welfarist principles of distributive justice could then avoid the Dworkinian objections I have been outlining if they restricted attention to the distribution of ordinary welfare. *Restricted* welfarism might be viable even if *unrestricted* welfarism leads to paradox.

Sacconi, Faillo, and Ottone's account might appear to suggest a way forward. For they define conformist utility partly by reference to social welfare functions that are specified to be principles that govern the distribution of *material* utilities. That is, the social welfare functions that figure in their account are *restricted* welfarist principles that govern the distribution of material utility,

*not* overall utilities. There is a Dworkinian rationale for this. Suppose that T is an egalitarian principle that has been unanimously accepted in our society and suppose that it is widely conformed to. Suppose now that you and I have conformist preferences of different strengths such that I strongly desire to reciprocate conformity to T but you only have a weak desire to reciprocate conformity to T. In this case, I would be more likely to conform with T than you would, and my overall utility would be enhanced much more when I conform to T than yours would be enhanced when you conform to T. Since I would gain extra welfare by my conformity to T, it follows that if T governed the distribution of overall utility, then other things being equal, T would require that you be compensated by giving you added welfare. That is, T would require me and others to compensate you by ensuring that your ‘material desires’ are more fully satisfied than would be required if you and I had conformist desires of the same strength. This seems manifestly unfair. An egalitarian principle should not require giving more in material utility to those who have a weaker desire to see equality achieved than to those who have a stronger desire. Sacconi, Faillo, and Ottone’s formulation avoids this difficulty. It bifurcates a person’s overall utility into a material component and a conformist component and it stipulates that a person’s conformist utility is enhanced only by her reciprocal conformity with a social welfare function that governs the distribution of *material* utilities in the relevant population. It identifies a set of restricted welfarist distributive principles that govern only the distribution of material utility. Call these ‘SFO principles’.

Unfortunately, Dworkin’s objections to welfarist principles are not fully avoided by Sacconi, Faillo, and Ottone’s proposed restriction to SFO principles. Let us look more closely.

The problem is that Sacconi, Faillo, and Ottone define conformist utility in terms of conformist preferences, but the distinction that they draw between conformist preferences and material preferences is artificial and delicate. Consider three examples. First, suppose I am an egalitarian who desires to see the fulfillment of a welfarist egalitarian program. My desire is not a desire to conform reciprocally to a principle that has been unanimously accepted. It is a desire for the existence of an egalitarian distribution. Hence it does not count as a conformist preference, on Sacconi, Faillo, and Ottone’s definition. By default, it must be a material preference. Second, even if I desire to conform reciprocally to a principle, my desire does not count as a conformist preference unless my desire is to conform reciprocally only on the condition that the principle has been accepted unanimously in the relevant population. So again, it must count as a material preference. Third, even if I desire to conform reciprocally to a distributive principle that has been accepted unanimously in the relevant population, my desire apparently does not qualify as a conformist preference unless the principle is a social welfare function that governs the distribution of material utilities. But suppose the principle I desire to conform with governs the distribution of resources in the society rather than welfare, or suppose it is Rawls’s difference principle that governs the distribution of ‘primary social goods’ rather than welfare (Rawls 1971). In these cases as well, my desire apparently does not

count as a conformist preference, and so must be counted, on Sacconi, Faillo, and Ottone's scheme, as a material preference. (For the definition of conformist preferences, see 293–5, esp. 294.)

Given what I have just argued, and given that conformist utility is defined in a specific technical way in terms of conformist preferences, the fact that conformist utility is not taken into account by SFO principles does not avoid Dworkin's objections. For, in Sacconi, Faillo, and Ottone's scheme, all utility derived from the satisfaction of desires that do not qualify as conformist preferences counts as *material* utility by default. This means that SFO principles are open to Dworkinian objections in each of the three kinds of example I discussed in the preceding paragraph. First, if I am an egalitarian who desires to see the fulfillment of a welfarist egalitarian program, the added utility I get when the egalitarian program is fulfilled does not qualify as conformist utility because, as I explained, my desire is not a conformist desire. Since this extra utility does not count as conformist utility it is included, by default, in material utility on Sacconi, Faillo, and Ottone's definition, and it can be taken into account by an SFO principle as enhancing my material utility. Second, suppose I desire to conform reciprocally to a principle, regardless of whether the principle is accepted unanimously in the relevant population, as long as it is widely accepted. In this case, since my desire is not a conformist desire on Sacconi, Faillo, and Ottone's definition, the extra utility that may come to me as a result of the satisfaction of this desire does not qualify as conformist utility. It counts as material utility by default, so it can be taken into account by an SFO principle as enhancing my material utility. Third, suppose I desire to conform reciprocally to a distributive principle that is not a social welfare function, but is instead a principle governing the distribution of resources or the distribution of primary social goods. In this case as well, as I explained, my desire does not count as a conformist desire, so the utility that may come to me as a result of its satisfaction does not qualify as conformist utility and it can be taken into account by an SFO principle as enhancing my material utility. SFO principles are open to Dworkinian objections in cases of each of these kinds.

Dworkin's objections arose for *unrestricted* welfarist distributional principles. It seemed that they could be avoided by *restricted* welfarist principles that govern the distribution of 'ordinary welfare' and that ignore the distribution of welfare that accrues to people due to the satisfaction of distributional preferences. The trouble is that in the kinds of cases I have described, a person's *distributional* preferences do not count as *conformist* preferences. As a result welfare that accrues to people with these preferences is counted as material welfare, and it can be taken into account by an SFO principle as enhancing their material utility. Consider, then, an SFO welfarist egalitarian principle that requires the equalization of people's material welfare. Suppose that I desire the fulfillment of this principle and you desire the opposite. These are not conformist preferences, as we have seen. Hence, as our society becomes more egalitarian, I gain material utility due to the increased degree to which my egalitarian desire is satisfied and you lose material utility due to the increased degree to which your opposing desire is frustrated. It follows that, other things being equal, the

principle in question would require compensating you and taxing me in order to correct for the added benefit to me of increasing equality. But an egalitarian principle should not require taxing egalitarians to compensate inegalitarians for the achieving of equality. This is the same objection as we saw before.

The objection here is a challenge to the coherence of formulating distributive principles as governing the distribution of material utility, on Sacconi, Faillo, and Ottone's technical definition of material utility, since material utility includes utility from the satisfaction of distributional desires. It is not enough to exclude conformist utility. To be sure, the objection is not an objection to Sacconi, Faillo, and Ottone's project, for it is not part of their project to answer Dworkin's objections to welfarist principles of distributive justice. The point, however, is that the reasons that led Sacconi, Faillo, and Ottone to define a set of restricted social welfare functions that govern only the distribution of material utility are of a piece with the arguments that Dworkin gives against unrestricted welfarist distributive principles. It is therefore worth noting that Sacconi, Faillo, and Ottone do not go far enough to avoid the Dworkian objections. A coherent welfarist distributive principle would have to exclude more than merely conformist utility. It would have to exclude utility and disutility that accrues to people in virtue of their distributional desires as well as their conformist desires.

It would have to exclude more than this. There are many desires the satisfaction of which would enhance my utility and the dissatisfaction of which would detract from my utility where it would be paradoxical or at least counterintuitive to take such utility to affect the justice of the distribution of welfare in society. For example, suppose that I strongly desire there to be water on Mars, and suppose there is water on Mars. This is an 'external' desire, as Dworkin might call it, since it concerns a state of the world that has no impact on my welfare as ordinarily understood. It would be counterintuitive for an egalitarian principle to count the extra utility I get as a result of the satisfaction of this desire as, say, upsetting equality so that I must be taxed to lower my net utility. It would be equally counterintuitive for a prioritarian principle to count this extra utility, say, as removing me from the group of worst-off and so as requiring, say, that I not be compensated for my poor physical health. Or suppose that I strongly desire the success of famine relief in Somalia. It would be counterintuitive to count the disutility I get as the result of the non-satisfaction of this desire as requiring that I be compensated. Surely it is the victims of the famine that ought to be helped, both in my eyes and in theirs.

It is unclear, however, exactly which desires are such that the utility one gets as a result of their satisfaction ought not to be taken into account by a principle of distributive justice. Indeed, it is arguable that welfare a person derives from the satisfaction of her external desires and distributional desires can sometimes be of concern on distributional grounds. Parents may have strong external and distributional desires regarding the welfare of their children. And it is arguable that the satisfaction or lack of satisfaction of such desires can enhance parents' welfare in a way that should be taken into account on distributional grounds. If you and I each have such desires for our children but if my desire is satisfied and yours is not, because, say, your child has suffered a terrible misfortune, it

seems that I am better off in a way that is significant on distributional grounds. Similarly it is not clear that conformist utility is never of concern on distributional grounds. Suppose that I am in a group that unanimously accepts rules for playing poker including a rule for distributing the winnings among members of the group, and suppose we reciprocally conform to the rules, getting a great deal of satisfaction from doing so. You, however, have not managed to find any such congenial group. It seems that in this situation I am better off as a result of the satisfaction of my conformist desire in a way that should be taken into account on distributional grounds.

Sacconi, Faillo, and Ottone have defined a class of restricted welfarist distributive principles, SFO principles, which govern only the distribution of material utility, not conformist utility. But in some cases it seems that utility due to the satisfaction of distributional and external desires also should be ignored on distributional grounds. And in some cases it seems that conformist utility should be taken into account on distributional grounds. I conclude then that although Sacconi, Faillo, and Ottone have made a start in addressing Dworkin's worries, more work needs to be done by welfarists.

#### 4. Conclusion

I have perhaps been unfair to Sacconi, Faillo, and Ottone. Their main concern, it seems, is to give empirical support to the view that, as they say, "economic agents are motivated both by consequentialist (and mainly self-interested) and 'conformist' preferences", which are desires "to act according to an agreed principle if complied with reciprocally by other interacting agents as well". Thus, they suggest, "the utility maximization model of a rational economic man" needs to be revised to represent these different kinds of desires as aspects of a "comprehensive utility function" (292). I have not disagreed with this basic idea. But I have suggested two things.

First, I do not see that Sacconi, Faillo, and Ottone's basic idea helps us to solve the compliance problem that faces the social contract tradition. The problem is to show that rational agents would comply with norms of justice or morality or, indeed, with any norms that require them to do otherwise than to maximize their expected utility. In Sacconi, Faillo, and Ottone's picture, conformist utility counts toward overall utility so the fact that agents can be moved by conformist utility does not show that they ever comply with norms that require them to do otherwise than to maximize their expected utility.

Second, I think that Sacconi, Faillo, and Ottone's bifurcation of overall utility into the two components of material utility and conformist utility is artificial. This is because, as I have shown, on Sacconi, Faillo, and Ottone's definition, the utility a person gains from the satisfaction of a variety of distributional and external desires counts as material utility even though these desires are not self-interested and even though these desires can be concerned with distribution in much the way that conformist preferences are. Some of these desires fail to count as conformist desires for minor technical reasons. So although it seems

obviously true that economic agents are motivated “both by consequentialist (and mainly self-interested)” desires and by desires that are external or distributional or conformist or principled and deontic, it is not clear that the latter class of desires is well-defined, and more important, it is broader than the class of conformist desires. Hence, if we are going to revise “the utility maximization model of a rational economic man” to take into account these different kinds of desires, we need to give more thought to defining the class of non-self-interested desires.

### **Bibliography**

- Dworkin, R. (2000), *Sovereign Virtue: The Theory and Practice of Equality*, Cambridge/MA
- Rawls, J. (1971), *A Theory of Justice*, Cambridge/MA
- Sacconi, L./M. Faillo/S. Ottone (2011), Contractarian Compliance and the ‘Sense of Justice’: A Behavioral Conformity Model and Its Experimental Support, *this issue*