University of Michigan Law School

University of Michigan Law School Scholarship Repository

Articles   Faculty Scholarship

1985

On Preferences and Promises: A Response to Harsanyi

Donald H. Regan
University of Michigan Law School, donregan@umich.edu

Available at: https://repository.law.umich.edu/articles/745

Follow this and additional works at: https://repository.law.umich.edu/articles

Part of the Jurisprudence Commons, and the Law and Philosophy Commons

Recommended Citation

On Preferences and Promises:
A Response to Harsanyi

Donald Regan

John C. Harsanyi sketches an entire normative and metaethical theory in under twenty pages.[1] Combining breadth and brevity, his essay is useful and interesting. It reveals the interrelations between Harsanyi's positions on various issues as no longer work or series of articles could do. But by virtue of its programmatic nature, the essay creates a dilemma for a commentator, at least for one who finds many things to disagree with. If I responded to Harsanyi in the same sweeping terms in which he argues, we would end up with little more than opposing assertions. At the other extreme, I could point out what seem to me particular defects in Harsanyi's arguments as they stand. But that would seem pedantic and ungenerous if my particular objections could be avoided by spelling out his arguments at greater length, as some of them surely could. (I am especially troubled by this possibility because I am persuaded that in the past I once misinterpreted an argument of Harsanyi's and treated it unjustly.)[2] Constraints of space prevent me from first spelling out Harsanyi's arguments as best I can and then criticizing them.

Accordingly, I shall limit myself to two topics, chosen partly because I think the observations I shall make are worth making whether or not they represent points of ineliminable disagreement between Harsanyi and myself.

I

Assuming that the basic point of morality is to bring it about that people's preferences are satisfied, Harsanyi says that "we cannot always use a person's observable preferences as our final criterion for his or her true interests" (p. 43). Rather, we should concern ourselves with people's true preferences, that is, with what they would prefer if they were fully informed and perfectly logical.

It is important to ask at this point just what sort of preferences Harsanyi is willing to dismiss as misguided. Is he taking the sort of

1. John C. Harsanyi, "Does Reason Tell Us What Moral Code to Follow and, Indeed, to Follow Any Moral Code at All?" in this issue; further citations to this essay will be in parentheses in the text.
2. See below, A Coda on Coordination.

position Richard Brandt takes?[3] Is he willing to throw out even such a fundamental preference as some adult's desire for artistic achievement, if, say, it can be shown that this preference was acquired as a result of childhood misperceptions about the means of or the importance of pleasing one's parents? Or is Harsanyi thinking only of instrumental preferences such as Jones's observable desire to go to Chicago, which is misguided because the reason Jones wants to go is to see his friend Brown, and Brown, unbeknownst to Jones, has moved to Los Angeles?

If Harsanyi contemplates ignoring as misguided only instrumental preferences, like Jones's desire to go to Chicago, then much of what I have to say will be irrelevant; but by the same token, his theory will be much less ambitious than it might seem, and it will require us to maximize the satisfaction of people's ultimate preferences, whatever they are and however much they may be historically based on ignorance, illogic, or psychological domination.

If, on the other hand, Harsanyi means, like Brandt, to subject even ultimate preferences to criticism by facts and logic—if he means to suggest that what is in a person's interest depends on what she would prefer after a course of cognitive psychotherapy—then I want to ask, Why? Why should we try to promote the satisfaction of people's "true" ultimate preferences, which are or may very well be hypothetical, instead of their observable ultimate preferences, which are actual? (Hereafter, whenever I speak of "preferences," I shall mean ultimate preferences.)

The question may seem an odd one. My purpose in asking the question is not to cast doubt on the importance of what distinguishes true preferences from observable preferences. I agree that it is desirable to be as well-informed as possible and to think logically in deliberating about practical questions. My purpose is rather to cast doubt on the ability of someone like Harsanyi, who takes the satisfaction of preferences as the fundamental nonmoral good, to account for the importance of information, logic, and deliberation—in particular to explain why information, logic, and deliberation matter in the formation of (ultimate) preferences, as opposed to the attempt to satisfy them.

To begin with, the only (ultimate) preferences I actually have are my observable preferences, not my true preferences if those are different. If my true preferences and my observable preferences diverge, what is the good to me of the occurrence of events that satisfy my true preferences but that disappoint my actual desires?

Faced with this question, Harsanyi might retreat one step. He might agree that, if a choice must be made between satisfying some agent's observable preferences and satisfying her true preferences, then it is her observable preferences that matter. But, he might add, the agent would be well-advised to deliberate about her preferences, or to get assistance

---

3. Richard B. Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979), esp. chap. 6.

in deliberation in the form of cognitive psychotherapy, so as to bring her observable preferences as nearly into line with her true preferences as possible. But again, Why?

The obvious answer is something along these lines: many of our preferences are misguided. They were encouraged historically by factual error, or they are the result of false ideals and inappropriate guilt feelings (or the like) acquired when we were under the sway of our parents. We have desires for things that give us no actual pleasure when we get them. And we have aversions to things that, if we were better informed or psychologically healthier, we would greatly enjoy. Deliberation will allow us to rid ourselves of these impediments to genuine happiness.

There are two problems with this answer. For one thing, it is by no means a necessary truth that we will enjoy life more if we discover and act on our true preferences. For example, it may turn out that our true preferences are harder to satisfy than our observable ones. Deliberation may destroy preferences we *can* satisfy (and the satisfaction of which would in fact give us pleasure) and replace them with preferences we cannot satisfy, condemning us to frustration.

Furthermore, the argument in favor of deliberation that I have suggested presupposes that the real good is not the satisfaction of our preferences, whether observable preferences or true preferences, but rather is the psychological enjoyment that, if we are fortunate, the satisfaction of our preferences produces. In other words, the argument for the superiority of true preferences presupposes that the real good is not preference satisfaction at all but is a particular sort of psychological experience. But it was precisely the inadequacy of a theory that posits only psychological experiences as the good that led philosophers to assert that the good was the satisfaction of preferences in the first place. If we abandon the psychological-experience theory of the good in favor of the preference-satisfaction theory, we cannot then turn around and use the quality of experience produced as the test for which preferences to satisfy.

In sum, deliberation about one's ultimate preferences is desirable— even, I should say, morally required—but the theorist who takes preference satisfaction as the fundamental good has no satisfactory explanation of why.

II

I turn now to a different topic, the relative merits of rule utilitarianism and act utilitarianism. One of Harsanyi's charges against act utilitarianism is that act utilitarians are unable to maintain many socially useful practices, most particularly practices like promising, which involve the creation of expectations (sec. 2).[4] I think Harsanyi is mistaken about this; act utilitarians

4. Harsanyi distinguishes between an expectation effect and an incentive effect of practices like promising. The former has to do with the degree to which people can form beliefs and feel secure and confident about what will happen in their future. The latter

can have a practice of promising. (Promising is the only practice I shall discuss, but I believe most proponents of act utilitarianism and many opponents would agree that promising can be taken as representative of a considerable range of similar practices.)

In my opinion, the issue of whether act utilitarians can influence expectations through a practice of promising is more complicated than Harsanyi makes it appear. I have discussed the matter at some length elsewhere, and I cannot discuss it fully here.[5] What I can do here, without repeating anything I have said elsewhere, is to suggest that Harsanyi misrepresents the institution of promising, not only as it would function among act utilitarians, but even as it functions in commonsense morality. In criticizing Harsanyi on these lines, I am not criticizing him alone. His view of how promising works is widespread among philosophers.

Imagine two friends, Damon and Pythias, who live in the same town and who see each other regularly. Damon and Pythias agree on Friday to meet for lunch the following Monday at their favorite Greek restaurant. Monday morning arrives, and Damon gets up late with a hangover. Not a terrible hangover, but enough of a hangover to make the prospect of the planned luncheon distinctly unappealing. Damon tries to telephone Pythias, without success. After brief reflection, he goes back to bed. He is breaking his promise, but he believes that he is justified in doing so under the circumstances.

Is Damon justified? I think he is, and I suggest that any commonsense moralist who was not engaged in writing a philosophical discourse on promising would agree with me. So, I believe, would Pythias. Most philosophers, however, write as if promises could be broken only when the costs of keeping them would be very great, and by that standard Damon would not be entitled to break his promise. Watching others eat while you have a headache, mild nausea, and no appetite for food is a hardship, but it is not "very great."

What would Harsanyi think about Damon's promise breaking? I take it there is no doubt that Damon and Pythias's agreement counts as an exchange of promises, whether or not they used the words "I promise." Harsanyi mentions agreements to meet one's friends at specified times and places as an example of the use of promising (sec. 2), and it is universally acknowledged that no particular form of words is necessary to making a promise.

Harsanyi says that commonsense morality allows a promisor to break his promise "only in some rare and rather exceptional cases where keeping

---

has to do with the degree to which people are willing to act on the basis of their expectations about how others will behave. See Harsanyi, sec. 2, and John C. Harsanyi, "Rule Utilitarianism and Decision Theory," *Erkenntnis* 11 (1977): 25–53, pp. 36–37. Since the incentive effect depends on the formation of expectations, I shall generally subsume both effects under the general rubric of expectations. I think there is only one point in my argument at which this conflation might matter (see n. 7 below).

5. Donald H. Regan, *Utilitarianism and Co-operation* (Oxford: Clarendon Press, 1980).

his promise would cause him or some other people extreme hardship, or where the promisee would suffer only a very minor loss if the promise made to him were not fulfilled" (p. 44). Certainly Damon's meeting Pythias as agreed would not impose extreme hardship. So the issue becomes whether Pythias will suffer "only a very minor loss" if Damon does not turn up. If the loss to Pythias is too large to count as "very minor," then on Harsanyi's test Damon would have to keep his promise; and if I am right that commonsense morality would allow Damon to break the promise, then Harsanyi's test would not capture the commonsense view. So it seems that the loss to Pythias here must count as very minor. But if this is a very minor loss, there will be many cases in which promisors are allowed to break their promises on the ground that the loss to the promisee will be very minor. And that is inconsistent with Harsanyi's claim that commonsense morality allows promises to be broken only in "some rare and rather exceptional cases."

It may seem that Harsanyi could avoid my present criticism by complicating his test just a bit. Instead of stating two excusing conditions in the alternative, one of which looks to the hardship to the promisor (or other people) and the other of which looks to the loss to the promisee, perhaps the test should require that the hardship to the promisor (or others) and the loss to the promisee be considered together and balanced against each other by some mechanism that embodies a presumption in favor of keeping the promise.

Certainly such a test would come closer to capturing commonsense promising, but it would not make Harsanyi's general view of promising tenable. The problem with his view is not that he fails to state a test that is perfect in detail. No one could do that. The problem, which came out in the last step of my discussion of his test as he states it, is his claim that commonsense morality allows promises to be broken only rarely. This, I think, is just not true. If I am right, then no test can capture commonsense promising which makes cases of justified promise breaking rare and exceptional.

Like most philosophers who write about promising, Harsanyi overlooks the fact that commonsense promising is a subtle and flexible practice. Some promises are more important than others. Some promises are meant by the parties to be taken more seriously than others. We cannot state necessary and sufficient excusing conditions in terms of a specific level of hardship to the promisor or a specific level of loss to the promisee. Promising does not work that way. (Although I have discussed only one example, it should suggest a host of others to the reader.) Furthermore, it is the act utilitarian who is best able to give a philosophical account of the subtleties of promising because it is the act utilitarian who is most sensitive to the importance of circumstances.[6]

6. Let me mention one other striking example of an overlooked subtlety in commonsense promising (recognizing that doing so takes me beyond the topics Harsanyi discusses). As I noted above, it is universally acknowledged that no particular form of words is necessary

Have I put too much weight on promises between friends or acquaintances? What about promises made in impersonal, commercial dealings? Surely those promises can be broken only rarely. I shall return to this point—I mention it now lest the reader become too impatient. But before I discuss it, there is more I want to say about Damon and Pythias.

It might seem that if act utilitarianism (or commonsense morality for that matter) were as permissive as I say about promise breaking, then act utilitarians (or commonsense moralists) would not be able to rely on each other's undertakings. I believe they can rely on each other. Imagine that Damon and Pythias are both act utilitarians. On Friday, they make their date for lunch on Monday. Each believes on Friday that, barring unforeseen developments, their meeting for lunch on Monday will be the best available pattern of joint behavior. (From this it follows, incidentally, that each believes on Friday that, barring unforeseen developments, the best thing for *him* to do on Monday, if the *other* keeps the date, will be to keep it also.) Now, Monday arrives. Can Pythias rely on Damon to turn up for lunch? Of course he can. To see this, it is only necessary to be clear about the relevant sense of "reliance."

If the question is whether Pythias can regard Damon's appearance as a cast-iron certainty, the answer is no. If the question is whether Pythias can assume that Damon will appear with a probability of 99 percent (or whatever probability would be guaranteed by most philosophers' overstringent account of commonsense promising), the answer is still no. But these are questions of little importance. The most important question is whether Pythias can reasonably assume that he ought to go to the restaurant himself. Pythias's going, in the expectation of meeting Damon, would constitute reliance on Damon's undertaking in one standard sense of "reliance," and this is the sense of reliance that matters most.

Should Pythias go? The obvious answer is yes. In order for it to be reasonable for Pythias to go, he does not have to know that Damon will certainly turn up, nor even that Damon will turn up with a probability of 99 percent. If the agreement was a sensible one when made on Friday, then in all likelihood Pythias should keep it as long as the probability of Damon's keeping it is, say, 50 percent or even somewhat less. We cannot be precise without more assumptions, but the point is clear: even if Damon will regard himself as excused by a minor hardship like his hangover, and even if Pythias knows that, the probability that Damon will turn up is still far greater than the minimum probability necessary to

---

for making a promise. Specifically, the words "I promise" are not necessary. But it does not follow at all that the precise words used are unimportant. In fact, use of the specific words "I promise" (or any of a few very close equivalents), while not necessary to the creation of a promise, normally increases the binding force of a promise. Use of those words constitutes an explicit acknowledgment by the promisor of the special importance to the promisee of this particular undertaking. There are many everyday contexts in which promises are made but in which use of the words "I promise," out of the blue, would be distinctly odd. Once we start thinking about just what words we use to make promises, and when and why, the nuances are legion.

justify Pythias in going to the restaurant—that is, to justify Pythias in "relying" in the sense that counts.[7]

There is still a puzzle here. I say that Pythias should keep the date if there is at least a 50 percent chance (roughly) that Damon will keep it. But on what ground do we expect Damon to keep the date with even that probability? Well, Damon will keep the date (probably) because he expects Pythias to appear. But that is true only because Damon expects Pythias to expect that he, Damon, will appear. And so on. We may wonder whether these mutual expectations can be reciprocally justifying, as it seems they must be for the agreement to work. As I say, this is a puzzle, but someone who argued on the basis of this puzzle that agreements between act utilitarians could not work would not be making the same objection that Harsanyi makes. Harsanyi's claim is apparently just that the act utilitarian promisor would be excused too often for it to be rational for the promisee to rely. I have argued that that is not necessarily true. If we accept the point I have made in response to Harsanyi's specific claim, then we can see that there may be two different equilibrium sets of mutual expectations between Damon and Pythias. If they have made no agreement, each one expects the other to turn up at the Greek restaurant on Monday with only whatever probability is suggested by that other's normal dispositions with regard to independent choices about where to eat. We can assume that in such a case Pythias does not expect Damon to go to the restaurant with enough probability so that it makes sense for Pythias to go there unless he independently feels like it. And vice versa. If, on the other hand, they *have* made an agreement, Pythias can expect Damon to go to the restaurant with a substantially higher probability than in the absence of the agreement (and vice versa), and this can be true despite the fact that each understands the other will not go to the restaurant if some unforeseen event makes going even a minor hardship. So, it can be rational for Pythias (for example) to go, barring an unforeseen event creating hardship for himself, in the expectation of meeting Damon, even though Pythias knows Damon will regard himself as excused by an intervening occurrence that makes it better on balance that he (Damon) not go given the *postagreement* level of probability that Pythias will go. In short, there are, as I have said, two different possible equilibria of (act-utilitarian) expectations. (There might be more than two, but that does not matter for now.) The puzzle mentioned at the beginning of this long

7. This is the point at which it might be important not to conflate the incentive effect and the expectation effect (see n. 4 above). Strictly speaking, what I argue for in the text is that Pythias will have sufficiently strong expectations to give him an adequate incentive to go to the restaurant. Even so, he may be slightly less confident of finding Damon there than he would be if he knew Damon was a rule-following promise keeper. Pythias may therefore lose some pleasure of anticipation. It seems to me that this slight loss in pleasure from anticipation (or, more generally, slight loss of confidence in one's predictions about one's own future) is likely to be outweighed by the case-by-case advantages that come from breaking promises when act utilitarianism says we should.

paragraph is just the puzzle of whether act utilitarians can move themselves from one such equilibrium to another by exchanging words, that is, by promising. Once we have got to this point, I think it is clear what we should say: it is an empirical question whether any particular group of act utilitarians have available to them a mechanism for moving between different equilibria of expectations; but since the expectations at both equilibria are completely act utilitarian, there is no logical reason why they should not possess such a mechanism.[8]

Earlier, I left a loose end, the question, Do I put too much weight on how promising works between friends or between people who at least have special reasons for mutual trust? As Harsanyi suggests, promises may be made in situations ranging from casual arrangements between friends to impersonal commercial dealings. Perhaps the act utilitarian's account of promising does not make promissory obligation firm enough to be relied on in trade or banking, at the other end of the spectrum from Damon and Pythias.

For present purposes, there are two principal differences between commercial transactions and dealings between friends. The less important difference is this: in a society in which the distribution of wealth does not conform to utilitarian principles, even a highly moral act utilitarian might sometimes attempt to secure a benefit for himself or another, in commerce, by an executory promise (a promise to be fulfilled at some future time) that he had no intention of fulfilling. This would not happen at all in a society organized completely according to utilitarian principles (nor, I assume, between friends in any society). It might happen, as I have said, in a society with a nonutilitarian distribution of wealth, but even here it would not happen often, for the simple reason that act utilitarians could not get away with this behavior very often without depriving themselves of the useful possibility of making promises they intended to keep.[9]

The much more important difference between commercial dealings and dealings between friends is this: if Damon is a good act utilitarian, then Pythias knows it. But Damon's banker, or Sears, Roebuck, or a supplier of raw materials to Damon's small business, does not know it. How, then, shall these commercial parties rely on Damon's promise?

There is a problem here, but it does not arise from Damon's being an act utilitarian. If Damon were instead a punctilious commonsense moral philosopher, who hardly ever broke a promise, how would Sears, Roebuck know it? They wouldn't. If commercial parties rely significantly on Damon's promises, it will normally be for one or more of the following reasons: (1) Damon deals repeatedly with a single promisee and thus has

8. For a parallel discussion, see Regan, pp. 32–43.
9. For my response to the argument that act utilitarians would be unable to take account of this loss of credibility because no individual false promise would have enough bad effects on act utilitarians' general credibility, see ibid., p. 67 and chap. 3.

a reputation to keep up. (2) Damon deals repeatedly with a class of promisees, of whom the particular promisee is one, who exchange information, so that Damon has a reputation to keep up. Or, (3) Damon's promise will be legally enforceable. All three of these reasons can apply even if Damon is an act utilitarian. (An act utilitarian can be bound by, and in certain contexts he can even support, laws that impose obligations without fully act-utilitarian excusing conditions. Of course, such laws would have greatly diminished importance in a society organized on thoroughly utilitarian principles.) With regard to relatively trivial transactions, commercial dealers may rely to some extent on the fact that most people keep most of their promises most of the time. But, as we have seen, that also will be true even of act utilitarians.

I hope I will not be taken as arguing that act utilitarianism and rule utilitarianism are really the same. I agree fully with Harsanyi that these are different theories, and I agree with him about what some of the differences are,[10] for all that I disagree with his claims about promising and similar institutions. Also, nothing I say here in defense of act utilitarianism should be taken to indicate a retreat from my belief that cooperative utilitarianism[11] is a better theory than either act or rule utilitarianism.

A CODA ON COORDINATION

One matter on which Harsanyi and I disagree fundamentally is the importance and complexity of the coordination problem. He dismisses the coordination problem as unimportant, and he thinks that in any event it is easy to show that rule utilitarianism does better than act utilitarianism at promoting coordination (n. 5 and sec. 5). I think he is mistaken on both counts, and I think both mistakes have a common source. Harsanyi gives an incomplete answer to a question that has long been recognized as one of the main stumbling blocks for rule utilitarianism: granted that it would be best if everyone followed rule utilitarianism, what should I do if some people do not follow it?

I do not suggest that Harsanyi overlooks this question. He addresses it both in his essay "Rule Utilitarianism and Decision Theory" and in his present essay. In "Rule Utilitarianism and Decision Theory" it is partly the recognition that some people will not follow whatever moral theory we decide is best that prompts Harsanyi's distinction between "flexible" and "rigid" agents.[12] If we adopt rule utilitarianism, then the flexible agents are rule utilitarians, and each flexible agent selects and acts on the best joint strategy for all flexible agents given the behavior of the rigid agents. In the present essay, the flexible agents are in effect the

---

10. Subject to what I say below in the Coda to this response, I agree with essentially everything in Harsanyi, "Rule Utilitarianism and Decision Theory," about the coordination effect (in Harsanyi's sense, i.e., as distinguished from the expectation and incentive effects).

11. Regan, chaps. 8–10.

12. Harsanyi, "Rule Utilitarianism and Decision Theory," pp. 34–35, 46–48.

people Harsanyi refers to as "all rational and morally motivated people" when he says the best code is the code that would maximize expected social utility when followed by those people (p. 44).

So far so good. But there is a further problem. How is one flexible agent to identify the other flexible agents? This is important since the best thing for the flexible agents to do may depend on how many of them there are. As far as I can see, Harsanyi just takes it for granted in both his essays that everyone knows who is who—who is flexible and who is not.[13] But the individual agent making a moral decision in the way Harsanyi recommends (if he does not jump over the whole process and simply assume that his society's existing moral code is the right one) must decide for himself who is flexible and who is not. As a practical matter, he will not do this before every decision, but the question does arise in principle in connection with any decision.

This creates a further problem. Suppose we accept tentatively that the best theory for the flexible agents to follow is rule utilitarianism. The first thing an individual rule utilitarian must do, then, is to figure out who else is a rule utilitarian. Note that this must be the precise question he asks about other agents. No more general inquiry will do. It will not suffice, for example, to figure out who else is a person of good will in some broad sense, such as, who is rational by ordinary commonsense standards and generally benevolent. The reason is that somebody might be rational by ordinary commonsense standards and generally benevolent and still be an act utilitarian, a follower of W. D. Ross, or a believer in some religious ethical code. But the rule utilitarian cannot count on such people to join him in selecting and applying the best joint strategy for him and them (as flexible agents), given the behavior of everyone else (say, the egoists and people of unusually weak will).

So, the first thing the individual rule utilitarian must do is identify the other rule utilitarians. But that means that the instructions to him are in fact a bit more complicated than merely, "Be a rule utilitarian." They are, "First figure out who the rule utilitarians are; then be a rule utilitarian." And, of course, it is important that the people the individual proposes to coordinate with (the other flexible agents) are also going through the same preliminary step, lest *they* make some false assumption about who is a rule utilitarian and is therefore flexible in the required sense.

If we call ordinary rule utilitarianism R, we have seen that the individual cannot merely follow R. He must follow R′, which says, "Figure out who

13. Perhaps what Harsanyi takes for granted is not that everyone knows who's who but, rather, that every flexible agent has expectations about who should be treated as flexible and who should be treated as rigid (see Harsanyi, "Rule Utilitarianism and Decision Theory," p. 47). In some contexts it would matter whether one assumed knowledge or only expectations, but I don't think it does for present purposes, since what I go on to discuss in the text is the question how the relevant expectations could rationally be formed and indeed whether there is any way they could be rationally formed.

else is following R; then follow R with them." Or rather, since the others must also go through the preliminary step, what R′ must actually say is, "Figure out who else is following R′; then follow R with them." Unfortunately, the instructions that are supposed to constitute R′ now contain within them a reference to R′. Can this self-reference be avoided? Can we spell out what the individual agent must do, including what he must do to identify the other agents it is sensible for him to try to coordinate with, in such a way that there is no vicious regress?

It was in demonstrating that we can provide an appropriate set of instructions, at least in principle, that I produced two very convoluted chapters in my book, which assume what Harsanyi correctly characterizes as "completely unreal conditions" (n. 5). I still regard it as worthwhile to have shown that the problem is solvable in principle. Harsanyi may of course disagree. In any event, those chapters were the best I could do by way of dealing with what we might call the "identification problem"— the problem of how to build into our instructions to an individual who wants to cooperate the essential first step of deciding who else is cooperating. That is a genuine problem, which I believe Harsanyi does not address.

Although Harsanyi does not address this problem, I said in *Utilitarianism and Co-operation* that he had come closer to addressing it than anyone before him,[14] and I meant that to be a favorable comment. The lucidity of his discussion of the coordination problem, *prescinding from* the identification problem, highlighted much better than any previous discussion just what remained to be done. However, I went on to do Harsanyi an injustice by treating him as if he also meant to deal with the identification problem. Assuming (falsely, I now believe) that he meant to address the identification problem (and to solve it by identifying the rigid agents with those whose behavior could be predicted), I attempted to explain briefly why his suggestions did not solve the problem (hardly surprising, if he was not trying to). In the process I made some arguments that are misguided and unjust if they are taken as criticisms of Harsanyi's treatment of the coordination problem (identification aside).[15] I apologize.

To return very briefly to my observations at the beginning of this coda, I suspect there are two reasons why I regard the coordination problem as more important than Harsanyi does. One reason is that, treating the identification problem as part of the coordination problem, I see many more tacit coordination problems than he does.[16] Voting and tax paying, indeed all matters involving obedience or disobedience to some general statute or regulation, present coordination problems that are "tacit" in the sense that they are not fully solved even by explicit agreement or legislation. In principle, the identification problem always remains. It is simply not susceptible of solution by fiat, whether it be the

14. Regan, p. 260, n. 7.
15. Ibid.
16. He says they are "rather rare and unimportant" (p. 48).

fiat of agreeing parties or the fiat of a legislator. Even in the simplest case, where two parties have made a face-to-face agreement, the question remains for each, Will the other follow it? In that sense, every coordination problem has an ineliminable tacit aspect.[17]

The other reason I think the coordination problem is more important than Harsanyi does is that I do not think the coordination effect and the expectation and incentive effects can be separated as completely as Harsanyi tries to separate them. Believing that act utilitarians can influence expectations and incentives by behavior such as promising, I also believe that for act utilitarians the problem of generating desirable expectations and incentives is itself a coordination problem and an important one. Moreover, if non–rule utilitarians must identify rule utilitarians and infer what rule utilitarians really believe at least in part from the rule utilitarians' *behavior*, then it seems to me that the problem of generating expectations and incentives is a coordination problem for rule utilitarians also.

17. I mention the problem of obedience to statutory and regulatory schemes only because almost every such scheme points to some coordination problem. It might seem that coordination is guaranteed in such cases without any specific thought by any individual of the desirability of coordination, since each individual is given adequate incentives for correct behavior by the threat of legal penalties. But the law could not function in the face of really widespread disobedience. The majority of people keep the system going by co-operating in obedience; and the penalties deter the remainder (insofar as they are effective).