GAME THEORY AND THE VIRTUES: THE NEW AND IMPROVED NARROWLY COMPLIANT DISPOSITION*

Grant A. Brown

University of Lethbridge

Two approaches to moral philosophy could hardly be more different than ancient virtue ethics and contemporary contractarianism. The former is abundant in its assumptions about human nature; it emphasizes historical continuities, particularized contexts, and "ordinary language;" it embraces a highly intuitive mode of drawing conclusions. The latter, by contrast, is austere in its assumptions about human nature; it is atemporal, non-contextual, and utilizes a specialized, "high-tech" vocabulary; it purports to be mathematically rigorous. As a (modern) paradigm of the former, consider Alasdair MacIntyre's *After Virtue.*[1] And as a paradigm of the latter, consider David Gauthier's *Morals By Agreement* (henceforth MBA).[2] Yet, in spite of the radical differences between these two types of moral theorizing, I believe that each could benefit by accommodating the strengths of the other. Game theory, the essential tool of contemporary contractarianism, can be used to tighten up virtue ethics, just as an appreciation of the traditional virtues can suggest

fruitful avenues for game theory to explore. I have criticized MacIntyre's approach as being too loose and indeterminate elsewhere.[3] Here I would like to subject Gauthier's views to criticism to show how some of the more traditional virtues which he ignores can be given a game-theoretic rationale.[4]

Gauthier advocates "narrow compliance." By this he means the disposition to respect free market rights whenever interaction is parametric; and whenever interaction is strategic, to co-operate with only those agents who in turn co-operate only in ways which yield nearly optimal and fair outcomes.[5] These outcomes are defined according to Gauthier's principle of minimax relative concession (MRC). This principle requires a distribution of the proceeds of co-operation in such a way that the largest concession any co-operator makes, relative to his maximal possible gain from co-operation, is as small a relative concession as is possible for anyone to make.

In this paper I argue that Gauthier misidentifies the *moral content* of the narrowly compliant disposition. Narrow compliance, as just specified, is neither a necessary nor a sufficient condition for rational social interaction, even among individuals who do not care about each other's interests; moral dispositions which go beyond this conception of narrow compliance are rational. In defending this claim, I extend Gauthier's mode of argument to some more traditional moral dispositions (e.g. reciprocal altruism, forgiveness, fortitude, moderation, and broad-mindedness), concluding that moral life is far more complicated than Gauthier represents it as being. Further, these complications pose serious problems for his demonstration of the strict rationality of narrow compliance. A weaker conclusion is indicated, such as that the common-sense institution of morality is not unreasonable.

The "Archimedean lever" by which Gauthier hopes to move the moral world is social ostracism only. If you are not narrowly compliant (if, for example, you are not co-operative enough because you maximize utility without constraint, or if, on the other hand, you are too co-operative because you interact with people who co-operate on terms less favorable to themselves than MRC), then it will be rational for other members of society to deny you the benefits of social interaction. In the long run you will lose more by this denial than you can hope to gain through not being narrowly compliant. Or so Gauthier claims. But obvious exceptions are ready to hand: imagine refusing to commission a life-saving work for a Mozart or a Marilyn Monroe simply because they co-operate with others on terms more generous than MRC! Some people possess special non-moral characteristics (e.g. genius, beauty, a sense of humor; in general "talents") which it would be more costly for some members of society to ignore than to cater to.[6] It is not rational to be *too* moralistic, to interact *only* with morally impeccable people. But Gauthier's theory

would require us to be moralistic to the point of ignoring our own utility functions when contemplating *specific* interactions with talented individuals who are not narrow compliers, since co-operating with them will tend in the long run to unravel his ideal co-operative society. (This is the burden of his argument at MBA, pp. 178ff.)

In this respect, Gauthier's theory is excessively demanding in a moralistic way. Compare D.A.J. Richards' "principle of mutual love requiring that people should not show personal affection and love to others on the basis of arbitrary physical characteristics alone, but rather on the basis of traits of personality and character related to acting on moral principles."[7] Commenting on this passage, Bernard Williams says, "This righteous absurdity is no doubt to be traced to a feeling that love, even love based on 'arbitrary physical characteristics,' is something which has enough power and even authority to conflict badly with morality unless it can be brought within it from the beginning. . ." Although Gauthier defends "free affectivity," the right to choose one's own emotional ties, he nevertheless, like Richards, must suppose that the power and authority of *all* talents can be brought within morality from the beginning, if there are to be no conflicts between the demands of his narrow compliance and rationality. I doubt that this can be done by any of the arguments Gauthier allows himself. Special talents give their possessors a lever by which they can nudge the moral world in their favor though, of course, how far they can deviate from narrow compliance depends upon how much weight their talents give them in the calculus of interaction, and also upon how many special cases there are.

The plot thickens considerably when it comes to people who are narrow compliers viz-à-viz members of their own group, but are straightforward maximizers viz-à-viz other groups. It would generally be irrational for members of the discriminated against groups to interact with these discriminators; but is this true also of like-minded members of the discriminating group? Not obviously, particularly when the discriminating group is relatively large. If this is so, and if discriminating groups overlap in very complicated ways, as they do, then it becomes increasingly difficult to say just what morality (and rationality) requires on Gauthier's theory.

Cases involving special talents illustrate that the disposition of narrow compliance, as articulated by Gauthier, is not a necessary condition for rational social interaction. Opportunities for reciprocal altruism illustrate this in another way. It would seem that Gauthier's narrow compliers apply the principle of MRC religiously, to each separate co-operative interaction;[8] and they do so without taking an interest in anyone else's interests. The point I wish to argue now is that the "formal selfishness" of Gauthier's co-operators limits the benefits they can hope to gain from social interaction, relative to what they could obtain if they were to adopt more altruistic dispositions.

Consider cases in which one person could make a tiny concession, in terms of resources, in order to yield an enormous benefit to another. By conceding a mere $35, Ernest could confer an additional benefit of $147 upon Adelaide (MBA, pp. 138-9).[9] Now, Gauthier claims that such a concession would be irrational for Ernest to make, since it could never be utility maximizing for him unless coercively exacted. He explains, "...it would be irrational for an individual to dispose herself to voluntarily making unproductive transfers to others. An unproductive transfer brings no new goods into being and involves no exchange of existing goods; it simply redistributes some existing goods from one person to another. Thus it involves a utility cost for which no benefit is received, and a utility gain for which no service is provided"(MBA, p. 197). It would seem that the kind of transfers presently under consideration are "unproductive" in this sense.

However, they are not necessarily irrational. Although such concessions are not *directly* utility maximizing, they may in some circumstances be *indirectly* utility maximizing. Thus in a society of reciprocal altruists, Ernest could expect Adelaide (or anyone else) to return the favour of making a small concession in order to provide him with a large benefit when circumstances were reversed. This would, in the long run, secure greater benefits all around. (Indeed, this strategy is structurally similar to the solution suggested by Narveson in endnote 9.) Genuine reciprocal altruism, which is different from Gauthier's articulation of narrow compliance, is an indirectly utility-maximizing strategy in a society of reciprocal altruists.[10] The formal selfishness exhibited by Gauthier's co-operators commits agents to an inferior long-run strategy (namely MRC), at least in this limited range of cases. The rationality of MRC is limited by its own presuppositions.

A narrowly compliant person is one with a disposition to co-operate in ways that are *nearly* optimal. Gauthier interprets "nearly" in terms of the *relative* concession an agent makes. For Ernest to concede the $35 he would get according to MRC would be for him to make a *total* concession; and it would require *no* concession at all from Adelaide. This is as far as one can get from "nearly optimal" on Gauthier's reckoning. But if we interpret "nearly" in terms of the distribution of resources, it is still plausible to see a total concession by Ernest as "nearly optimal" he does not lose much, in terms of resources, in relation to what Gauthier would give him. My suggestion is that we should interpret the narrowly compliant disposition to include reciprocal altruism in cases in which it can be claimed that a total concession is "nearly optimal" in resource distributions, rather than relative concessions. In such cases, perhaps it could be said that the transfer is productive after all: productive of goodwill on the part of the altruist.

So far I have adduced considerations which tend to "broaden" the

allowable range of actions that can be rational under a narrowly
compliant disposition. The remaining considerations will introduce bases
of discrimination among narrow compliers as more broadly understood
above. It is not rational to interact *only* with narrow compliers, as
Gauthier understands this disposition; nor is it rational *always* to interact
with narrow compliers so understood even where there is a co-operative
surplus to be realized. Making this latter point requires some setting up.

The official version of Gauthier's theory sets each person's initial claim
at his *maximum utility level* (MBA, p. 134). Now, most of his examples
are very simple, typically involving dollar returns; and in discussing these
examples, Gauthier makes the always-dangerous simplifying assumption
that peoples' utilities are linear with monetary values (MBA, p. 137). But
the assumption that people's utilities are linear with the quantity of any
particular good is entirely unrealistic. Market theory is based on the idea
that declining marginal utilities for goods make trade opportunities
abundant. In the ensuing paragraphs, I would like to pursue a more
complicated and admittedly fanciful case where the dangerous simplifying
assumptions used by Gauthier are relaxed. It is fanciful only because I use
a single example to illustrate a number of distinct points, thereby making
it extreme.

Here is the scenario: Two children, Veronica and her brother Norm, are
given, jointly, ten hours of television viewing time per week, provided that
they complete various household tasks. That is, Veronica and Norm must
complete a joint venture (household tasks) in order to realize a
co-operative surplus (television viewing time). Furthermore, they must
decide in advance how they will accomplish this joint venture (i.e. divide
up the tasks), and also how they will distribute the surplus.[11] The relevant
consideration from the point of view of Gauthier's theory of rational
bargaining is the participants' utility functions, so it will help to
appreciate the difficulties of the case if we characterize Veronica and
Norm a bit more fully. As far as the tasks the costs of the joint strategy
go, then, it will be important to note that Norm has a rather typical
displeasure threshold. That is, he can tolerate the everyday tasks of life
with (near) equanimity. Not so Veronica. She has a very low displeasure
threshold, and finds even the most mundane tasks rather taxing upon her
patience. Awareness of this throws her into the deepest misery, from
which only watching T.V. can rescue her.

Turning now to the benefits, we should note several features of our two
protagonists' psychologies. For the most part, Norm has a typical utility
function involving declining marginal utilities: he derives most of his
viewing pleasure from the first hours of T.V. watching, and steadily less
and less the more he watches (but always getting some positive
satisfaction therefrom, at least up to the ten hour limit). Veronica, on the
other hand, is a "resource monster": she cannot get enough T.V. viewing,

and each additional hour adds at least as much pleasure as the first. In fact, her pleasure seems to feed upon itself; the awareness of having greater pleasure increases her pleasure still more.

Veronica has expensive tastes as far as the T.V. goes. Above all else, she prefers mini-series, which spread themselves out over five or more one-hour segments. Of course, most of the utility she experiences is derived from the final episode, when the plot of the series is resolved; but she cannot miss any of the prior episodes without losing the benefits of the whole series. Norm, by contrast, is happy to watch a disconnected series of one-hour or even half-hour programs; he tends to get rather bored if a show drags itself out for too long.

Norm has reasonably broad tastes in television shows, and can watch comedy, sports, news and documentary, nature and travel, or various other types of programing with almost equal pleasure. He does have distinct preferences, but these preferences are overridable. Thus he would generally prefer to watch two hours of Veronica's most preferred shows to only one hour of his own most preferred show. Veronica, however, is a fanatic. She can be satisfied only by the show she prefers most at any given time. Thus if she had to sit through one of Norm's shows, she would be all but indifferent; she would contemplate the (relative) "loss" she is "suffering" as much as the benefit she is receiving.

Veronica has a best friend, Monique Jones. Monique is an only child, and has her own T.V. Veronica and Monique like nothing better than to chat about their favorite T.V. personalities and shows (which guy is the cutest, what's going to happen next episode, and so on). When Veronica cannot see everything that Monique sees, they are unable to chat as successfully as before, and this displeases her. It also pains her that she is unable to "keep up with the Jones's" in terms of T.V. viewing time, which is something of a status symbol in her circle of friends.[12] Norm and his best friend, on the other hand, rarely talk about T.V. shows. When they are together they create their own enjoyments in the form of playing games.

To summarize: Whereas Norm has a reasonable level of fortitude, Veronica is faint and delicate. Also, Norm's preferences are temperate, moderate, broad-minded, and non-competitive; Veronica is a fanatical and competitive resource monster with expensive tastes. Given these psychological profiles and MRC, we cannot determine whether Norm will end up doing most of the household tasks with a fairly even split of the T.V. viewing time, or whether they will split the tasks evenly with Veronica getting most of the T.V., or what. What we do know is that Veronica's share of the T.V. will be significantly greater than her share of the household tasks.

Intuitively, this seems unfair; one might even say that Veronica exploits Norm's humaneness. Yet Gauthier's explicit theory bars him from seeing

matters that way; he provides no rational basis for criticizing these distributions of tasks and benefits. I would like to suggest that this is a fault of his theory and not of our intuitions. Indeed, our intuitions here can be bolstered by a more careful and intricate application of game-theoretic reasoning. The point I am making is not simply that a utility-defined theory of justice can have bizarre resource-distribution consequences under suitable assumptions. What is special about the above example is that it illustrates various virtues (Norm's) and vices (Veronica's) which should be rationally taken into consideration when deciding upon the terms of interaction -- or when deciding to co-operate at all. Our common-sense intuitions about this case point to an inadequacy of Gauthier's reasoning.

Nothing in the story suggests that Veronica is not a narrow complier, as Gauthier understands this disposition. We may stipulate that she is one. Yet the existence in society of people like Veronica poses a problem for normal folks. Whenever co-operation is required of them,[13] they act like drains on the co-operative surplus in that they command a greater share of the benefits, in resource terms, while contributing less to bringing it about. This is a "public bad" which normal folk could do well without. If Veronica had had character traits more like Norm's, both she and Norm would have done better (in terms of utility) in their interaction. It would be irrational for anyone to interact with Veronica-like people if they had a choice. The vices exhibited by Veronica should be recognized as such, and as a matter of rational interpersonal policy they should be put in their place, not encouraged. Contrariwise, the virtues illustrated by Norm should by encouraged, not frustrated, by social interaction.[14]

What the case of Norm and Veronica illustrates is that preference structures themselves can have a "moral tone." That is, there are morally relevant features of people's preference structures which call for very discriminating responses. Just as Gauthier's constrained maximizers take positive account of the utility levels of those with whom they interact co-operatively (MBA, p. 167), so would my narrow compliers take into account the underlying psychological bases of these utility levels. Rational individuals would attempt to discriminate the characters of those with whom they interact in ways not anticipated by Gauthier. They would, so far as they were able, prevent the mean-spirited, spineless, and utility-consuming Veronicas in society from benefiting abnormally (in terms of resources) from co-operative interaction, just as they would, so far as they were able, exclude straightforward maximizers.[15] A careful application of game-theoretic analysis recommends this to rational people. This conclusion is not fundamentally antithetical to Gauthier's project, it merely extends to some more traditional moral dispositions (fortitude, temperance, moderation, broad-mindedness, etc.) the same rational basis on which Gauthier puts constrained maximization. In so

doing, I believe it makes the contractarian approach to morality much more rich, realistic, and attractive.

Many other virtues and vices can be given a solid game-theoretic rationale as well. An important virtue that deserves special consideration is forgiveness. It is widely acknowledged that Tit-for-Tat the strategy of repaying nonco-operation and co-operation in kind is the best means of securing co-operation in reiterative games in which optimality requires it. Indeed, Tit-for-Tat is an important component of the disposition of narrow compliance, as Gauthier understand it. What he neglects to mention is that when playing Tit-for-Tat, one must be prepared to switch to the co-operative strategy oneself if one's partner relents and shows himself willing to do so also. To hold a grudge after the first unco-operative move is to be locked into nonco-operation, which is suboptimal. Hence the rationality of a disposition to forgive.

Without going on to summarize the literature in this area, I wish to suggest that a new and improved narrowly compliant disposition which is in line with the arguments I have sketched above will not be unrecognizably distant from common-sense morality. (*Plus ça change, plus c'est la même chose.*) Perhaps surprisingly, Gauthier's contractarian-ism, suitably amplified, is very much compatible with a virtue-oriented ethics. This is less surprising if one attends to Gauthier's emphasis on the primacy of dispositions throughout his exposition. He identifies rationa-lity at the level of dispositions virtues and carries through the implications of this for individual acts.

I conclude this paper by suggesting, equally sketchily, that contractarian morality, fully developed along these lines, cannot be demonstrated to be strictly rational. In my view, the most that can be said about the relation between morality and advantage is that, in general, on balance, and in the long run, if one is not too unlucky, these will not often clearly conflict. By this I do not mean that it is a toss-up between adopting the full panoply of moral dispositions on the one hand, and adopting no morality on the other. Rather, it seems to me that there is a "critical mass" of central moral dispositions which are rational requirements of any social interaction; but that beyond these rather minimal requirements, the advantage of specific moral dispositions cannot be strictly demonstrated in the abstract.

If the moral landscape is far more rich and complex than anything Gauthier's explicit theory indicates, this has serious consequences for his arguments for compliance. Indeed, the simplicity of the situation facing Gauthier's moral agents is crucial to his demonstration of the strict rationality of narrow compliance. Gauthier introduces only two compli-cating factors: that people's dispositions are not completely transparent, and that the general population contains people who practice a mixture of co-operative and nonco-operative strategies (MBA, pp. 174-79). Yet

even these complications force him to qualify seriously his endorsement of constrained maximization. Once we realize that a person can instantiate only a small selection of the wide array of legitimate moral characteristics, each only to a greater or lesser degree, it becomes evidently impossible to distinguish sharply between the sheep and the wolves. Most people inhabit the vast grey area in between. How, then, can ostracism work as a reliable means of shaping social interaction? The Archimedean lever by which Gauthier hopes to move the moral world is in fact rather pliant.

It is highly doubtful that quasi-mathematical calculations will produce compelling results once all of the relevant considerations are factored into the equations. There is no convincing way to carry out a large scale cost-benefit analysis which takes account of all the necessary variables such things as interaction with morally imperfect people who possess special talents, reciprocal altruism, forgiveness, and discrimination of various virtues and vices. In fact, I do not believe that many very specific moral principles, beyond rights to personal security and obligations to honor one's word, can be formally demonstrated to be rational requirements of all social interaction. Morality, it seems to me, is underdetermined by formal, game-theoretic rationality, which is precisely why we must depend upon "practical reason" a more intuitive, contextual mode of appraising moral situations.

Philosophers of science have come to realize that even our most central theories are rationally underdetermined, yet we need not follow Descartes and be frightfully concerned about this. The same is true of morality. What moral theory must do is provide a schedule of values, roughly ordered in terms of centrality and stringency, leaving each society to give shape to these values in their concrete social and political processes. What is not possible, what we should try to avoid, is to derive from pure reason a very fine-grained systematization of moral values, applicable to all societies. Morality is indeed "made" or constructed by agreements, not by philosophers.[16]
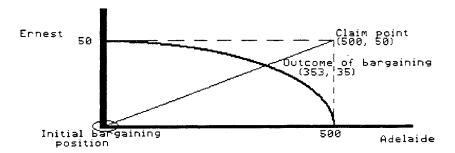
1. Alasdair MacIntyre, *After Virtue*, 2nd ed., (Notre Dame: University of Notre Dame Press, 1985).
2. David Gauthier, *Morals By Agreement*, (Oxford: Clarendon Press, 1986), p. 178. 3.

Grant Brown, Review of *After Virtue*, *Eidos*, Vol. VII, No. 1 (June 1988): pp. 105-10. A footnote in which I suggested how game theory might be used to improve MacIntyre's position was omitted in the published version.

4.   In the preface to *Morals By Agreement* Gauthier says, "...the conception of practical rationality that I accept at the root of my argument [i.e. the instrumental, maximizing conception which is foundational to game theory] seems to me the only one capable of withstanding critical examination, and the moral theory that I then develop seems to me, in outline if not in every detail, the only one compatible with that conception of rationality. Yet, as Richard Rorty or Alasdair MacIntyre might remind me, perhaps I lack the vocabulary for talking perspicuously about morality." (p. vi) Here I argue that when we properly work out the details of morality given the approach Gauthier sketches, we come to appreciate how the vocabulary favoured by Rorty and MacIntyre can be accommodated within that approach. Elsewhere I have argued that Gauthier's "constrained maximization" is really a satisficing disposition, as are many of the virtues we intuitively recognize (e.g. moderation and spontaneity). See Grant Brown, "Satisficing Rationality: In Praise of Folly," *Journal of Value Inquiry*, in press.

5.   Interaction is parametric when the actions chosen by one agent do not affect the payoffs to the other agents. Each agent can therefore regard the payoffs being faced as fixed by the circumstances, with only the agent's own choice of action being a variable. Interaction is strategic when the payoffs to agents depend upon what other agents do. Thus there are interdependencies in strategic interaction, and agents cannot determine their payoffs given only their own actions. The most explored kind of strategic situation is the Prisoners' Dilemma.

6.   Gauthier assumes that all members of society are equally and fully rational, but he is not entitled to assume that everyone is equally talented.

7.   The principle is advanced by D.A.J. Richards in *A Theory of Reasons for Action*, (Oxford: Oxford University Press, 1971), p. 94. It, as well as the quotation following in the text, is found in Bernard Williams, *Moral Luck*, (Cambridge: Cambridge University Press, 1981), p. 16.

8.   As will be noted presently, there is some tension in Gauthier's exposition. In a later article Gauthier notes, "I do not attempt to deal with the relation between macrolevel fulfillment of MRC and microlevel principles of interaction." (Cf. "Moral Artifice". *Canadian Journal of Philosophy* 18 (1988): p. 390. That is, a society is just, according to Gauthier, as long as the interaction of its members roughly satisfies MRC on the macrolevel, regardless of the microlevel principles employed to bring this about. Here I argue that those "microlevel principles of interaction" diverge considerably from MRC.

9.   Below is a geometric representation of the bargaining situation.

Jan Narveson notes that in cases where outcomes can be assigned a transferable value, the indicated solution is *always* to opt for the joint strategy that produces the greatest value (in the present example, Adelaide's way for $500), and then compensate the party who would otherwise lose out in this strategy (Ernest) with a side-payment (in excess of $35, but less than $147). This solution dominates MRC. Though valid and important in other contexts, I will accept Gauthier's verbally communicated response to this criticism, which is to say that we are concerned here with the *logic* of the case for MRC, and so will assume that the goods in question cannot be transferred. (But doesn't Gauthier's normative theory of rational choice imply that *all* goods can be assigned a transferable value indeed, a monetary value? Cf. the discussion of the requirement that preferences be continuous (MBA, pp. 45-6).)

10. R.L. Trivers maintains that reciprocal altruism is the basis for many evolutionarily stable strategies found in nature. Gauthier's conception of narrow compliance might also be a stable strategy, as he suggests in citing Trivers (p. 187); but I think it is important to emphasize that a disposition of narrow compliance is not as comprehensive as reciprocal altruism.

11. As I construct the case, it is a non-market, strategic interaction. That is, Norm and Veronica cannot shop around for more congenial partners who may be willing to interact on better terms. Thus the terms of the joint venture are not determined by the market and individual rights, but by bargaining. This may seem artificial; however, like Gauthier, I am only concerned here to illustrate the logical implications of MRC.

12. Veronica is a "positional goods" seeker. As such her utility function is not completely independent of others' utility functions. But her interactions with Norm still exhibit mutual unconcern, or non-tuism (MBA, p. 87).

13. That is, whenever interaction is strategic. An n-person example which is very much in vogue these days concerns the co-operation needed not to avoid polluting our collective environment beyond recovery.

14. Paul Viminitz suggests that societies will tend to gravitate toward either Norm-like or Veronica-like dispositions, depending upon the relative proportions of each initially existing in the population. This is because it is easier for each individual to change his or her own characteristics than it is to effect change in most everyone else's. But for reasons given below, having to do with the inconclusiveness of compliance arguments, I think that a certain proportion of Veronicas could be "evolutionarily stable" within a larger population of Norms. This is why I believe active attempts to discourage those dispositions are reasonable.

15. Another possible response to the problem of Veronicas would be to insist upon an equitable division of the costs and benefits of social co-operation, regardless of the initial claims advanceable by the different agents. This is a less desirable solution in individual interaction because it is less discriminating. But in real-life, n-person situations, the bargaining costs associated with other solutions may make the equitable one a salient and (therefore) optimal one to pick. As Mike Kubara impressed upon me, we must never forget to take into account the costs of insisting upon precise justice, which in many cases are considerable.

16. Cf. Gilbert Harman, "Justice and Moral Bargaining;" *Social Philosophy and Policy* 1 (1984): 114-31. Sociobiology provides a useful model here, by illustrating how optimal behaviour patterns (whether genetically programmed or learned) are highly sensitive to variations in the local environment and to initial conditions. We should expect our "microlevel principles of interaction" to be likewise sensitive to these social variables.