

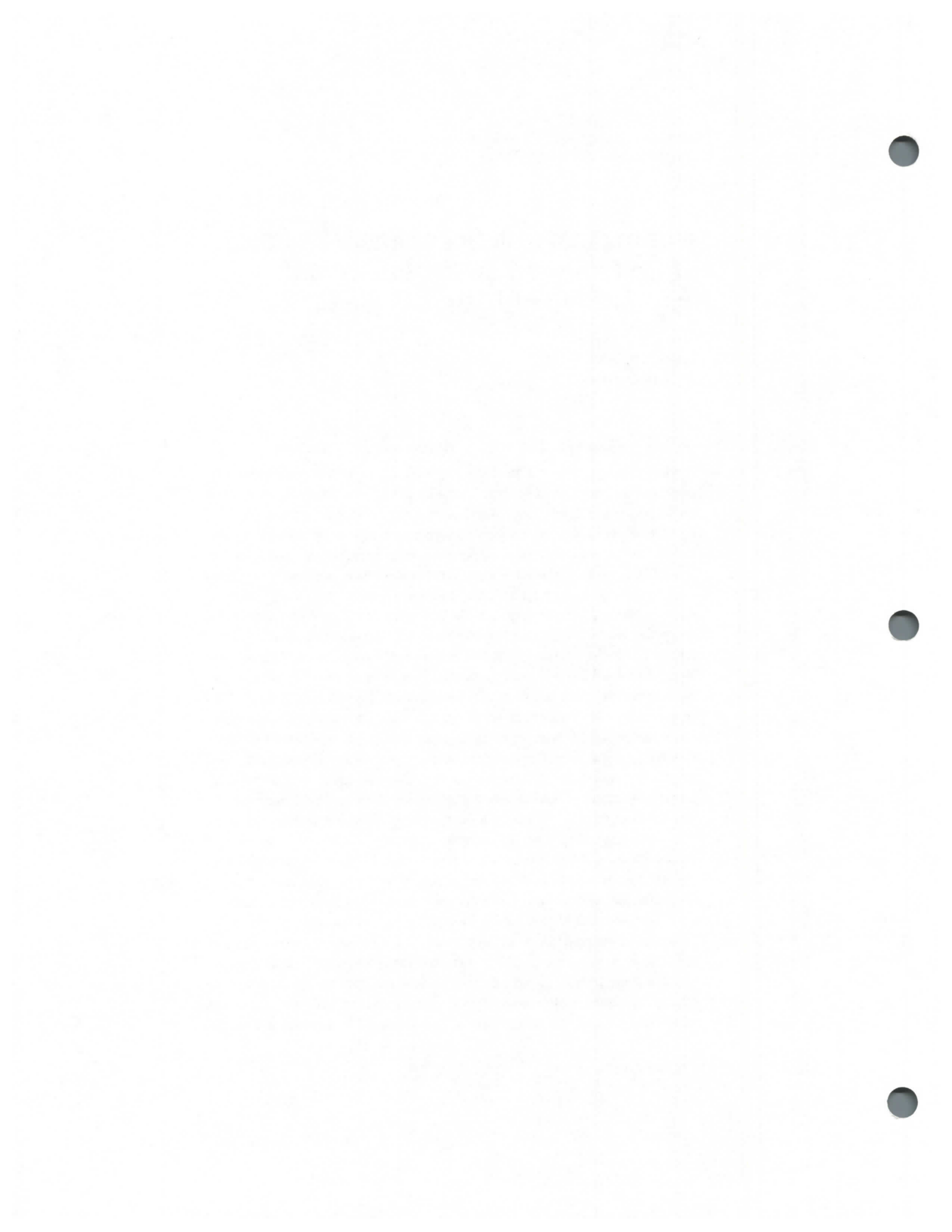
Commentary on Sober and Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior*

DANIEL C. DENNETT
Tufts University

Altruists, Chumps, and Inconstant Pluralists

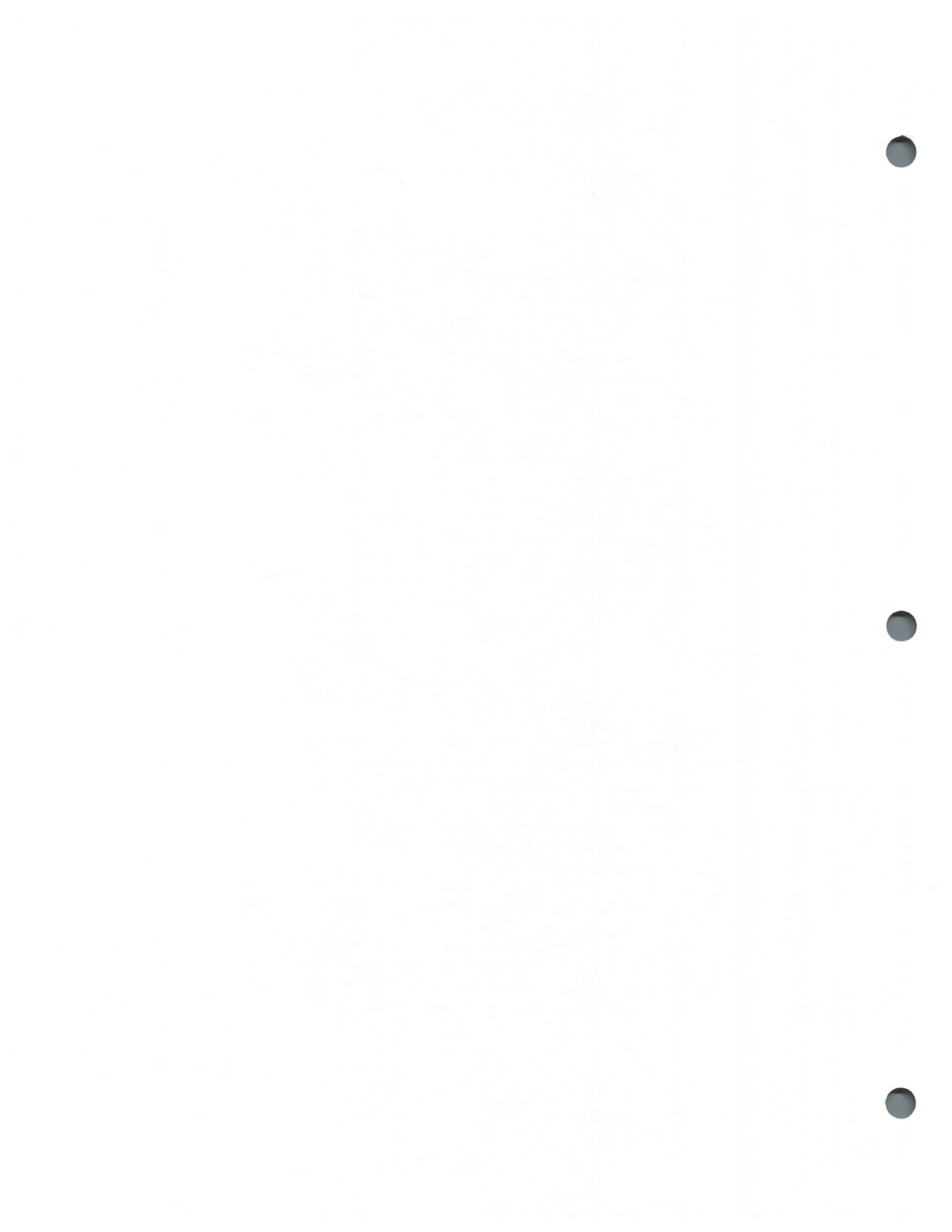
Anybody interested in evolutionary explanations of social phenomena (and every philosopher should be) will learn a lot from *Unto Others*. In addition to its cornucopia of fascinating empirical findings from biology and psychology, it is chock full of arresting perspectives, ingenious thought experiments, and clear expositions of difficult—indeed, treacherous—concepts that should be in every philosopher's kit. What philosophers will not learn, however, is the status of group selection in current evolutionary theory, because while Sober and Wilson (hereafter S&W) strive intelligently to clarify and unify the issues, some of their efforts muddy the waters instead. This is mainly due to the lingering effects of ancient feuds and score-settling on both sides. So far as I can see, nobody has yet been able to come up with a judicious overview of the whole scene. Evolutionists love to pick a fight. Why propose a friendly amendment when you can claim to overthrow some major edifice of orthodoxy with your revolutionary idea? And what could be more satisfying than resurrecting a derided heresy? Stephen Jay Gould briefly tried to restore Goldschmidt's "hopeful monster" saltationism as part of his campaign for punctuated equilibrium, which was itself not as revolutionary as he claimed (Dennett, 1995); hardly a year goes by without somebody declaring that something they have discovered is, actually, a genuine instance of Lamarckianism. There is usually a grain of truth, or more, in these dramatic claims. Prions really do constitute a major exception to Crick's Central Dogma about the direction of information flow between DNA and proteins

Have Sober and Wilson salvaged a sophisticated and sound perspective for group selection from the rhetorical overkill of the selfish-gene's-eye gang, or have they merely reinvented (or just relabeled) Hamilton's and Maynard Smith's *alternative* to group selection models, models that can do justice to all the observed and even imagined phenomena of cooperation in the biosphere? One of the main lessons I have learned in thinking about the



issues raised by *Unto Others* over the last two years is that they are, at least for me, mind-twistingly elusive and slippery. The appeal of the competing metaphors is such that there is unwitting sleight of hand in every direction, as the perspectives shift back and forth. One stern admonition might then be: eschew the metaphors, stick to the math! The problem is that the math doesn't distinguish between the perspectives Sober and Wilson champion and those they oppose. By their own account, the mathematical models are equivalent in what they can predict—and hence equivalent in what they can, in *one* important sense, explain. But that is not the end of it: "When one theory achieves an insight by virtue of its perspective, the same insight can usually be explained in retrospect by the other theories. As long as the relationships among the theories are clearly understood, this kind of pluralism is a healthy part of science." (p. 98). Good point. Patrick Suppes used to delight in challenging cognitive psychologists to point to any "cognitive" phenomenon that he couldn't model, retrospectively, in strict behaviorist terms. The wise response was to deflect the challenge. His behaviorist models were parasitic: they would never have been devised—the phenomena in question were all but invisible to the behaviorists—without the inspiration of the cognitive model they translated into behaviorese. And what Sober and Wilson say here echoes what one of their chief opponents, Richard Dawkins, had already claimed for *his* perspective: "The extended phenotype may not constitute a testable hypothesis in itself, but it so far changes the way we see animals and plants that it may cause us to think of testable hypotheses that we would otherwise never have dreamed of." (*The Extended Phenotype*, 1982, p. 2) So what's all the fuss about? If each side is just saying "Here's a good way of thinking of things, a prolific generator of hypotheses to test!" they could both be right, with their different perspectives having rather different utilities in different circumstances.

If Sober and Wilson had settled for the pluralistic reading of their advertisement for the tactic of adopting "multilevel selection theory," all might be well, but at times their pluralism evaporates, replaced by charges of error and confusion—charges which have been angrily reflected back by the other side. In the interests of de-escalation, let me comment on some recent gossip I have heard, to the effect that the late and deeply lamented William Hamilton had declared himself in a letter to Wilson to have "gone over" to the side of the group selectionists! That is an unfortunate twisting of a peaceful gesture. Not long after *Unto Others* came out, Hamilton, Alan Grafen, Mark Ridley and I met at Richard Dawkins' house to discuss the issues, since I was persuaded, at the time, that S&W had indeed uncovered a deep oversight in the gene's-eye view, and wanted to try to explain to them what I thought it was. The upshot of that long and sympathetic meeting was that I learned otherwise: the assembled theorists did indeed agree with many of S&W's central

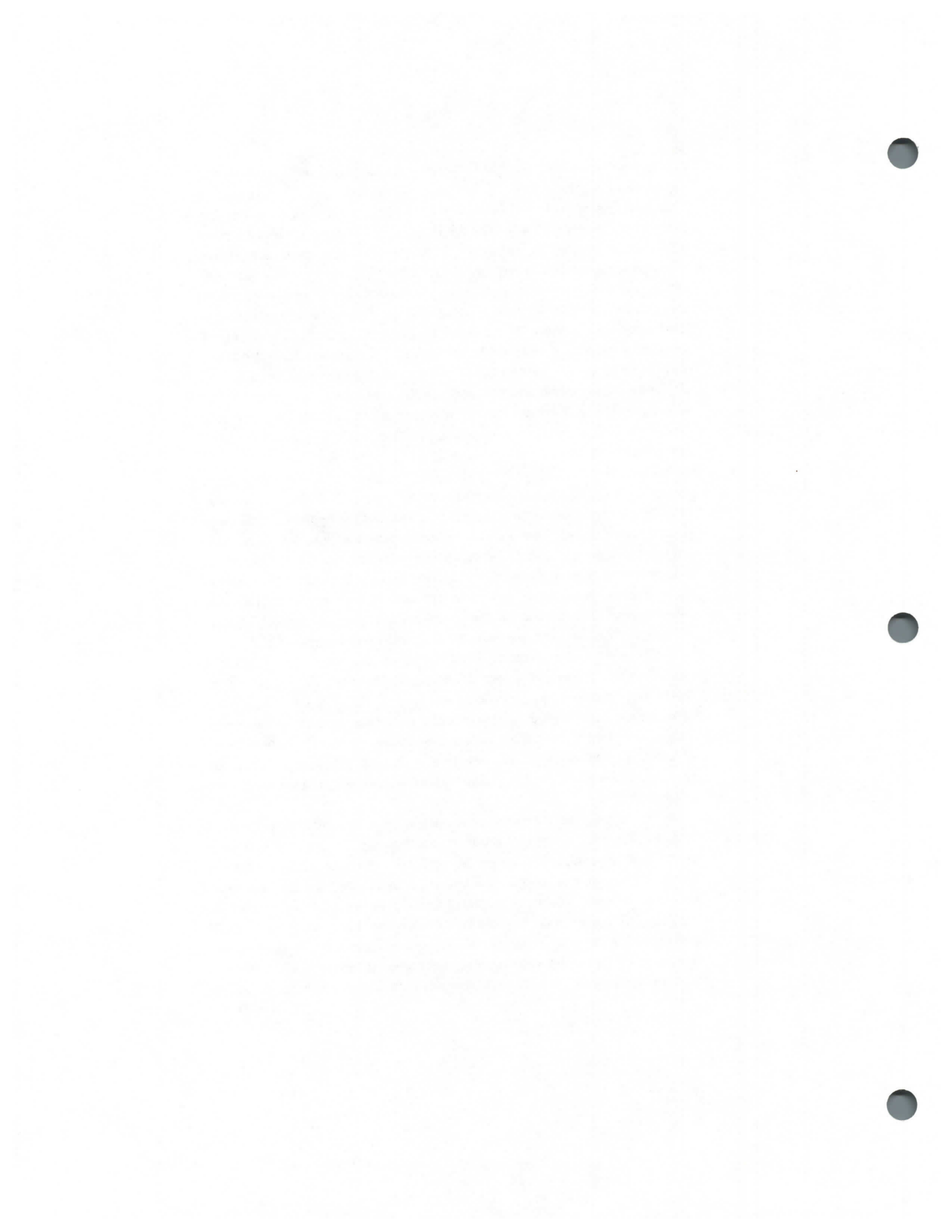


points, but claimed that these had all been quite explicitly anticipated, though expressed in different terms, by Hamilton thirty years ago. (See Hamilton's *Collected Papers*, Vol. 1, 1996, especially chapters 5 and 6, and his recollections of his "collusion" with George Price, pp 171-76, to get Price's paper published in *Nature*.) According to Sober (personal correspondence) Hamilton's subsequent letter to Wilson chided S&W for not including his work among the anticipations of multi-level selection theory. The two sides were much closer than I had appreciated—but also much closer than S&W portray them as being. Question: Do the Hamilton/Maynard Smith models predict all the effects S&W use to illustrate their multilevel selection theory? Answer: Yes. More pointedly, haven't S&W just shown that Maynard Smith's haystack model has a rather wider potential application than he envisaged? So far as I can see (and I've learned that my own perception of these matters can be readily tricked), that is what is going on. But that's enough, a valuable contribution.

S&W define (*evolutionary*) *altruism* as "behaviors that decrease relative fitness within groups but increase the fitness of groups," (p. 99) but this way of putting it conceals an important factor (acknowledged by them): "self-sacrifice on behalf of the group *can* evolve by group selection, but it is never advantageous *per se*." (p. 31). When, then, can it evolve? Only when, thanks to the particularities of the groups involved, those individuals whose relative fitness decrease *within* their groups nevertheless (thanks in part to their contributions and those of others in the group) enjoy a fitness *increase* relative to those in other groups not so blessed by the presence of altruists.

S&W object to the perspective that highlights this fact: "The existence of more than one group and fitness differences between the groups have been folded into the definition of individual selection, defining group selection out of existence." (p. 32) but their definition might be accused of the mirror image bias: "folding in" *the reason* for the increase in *overall* relative fitness of these altruists (being a member of a *successful* group) by calling the group's success an instance of *group* "fitness"—defining group selection *into* existence.

In one regard, what they propose is undeniably a valuable way of looking at some phenomena. It draws attention, as they say, to the role that *mere* group structure *can* play. Consider an artificial puzzle: we take 100 altruist clones, as like in their cooperative tendencies as can be, and sprinkle them around in different groups: some found thriving lineages, others go extinct. Why the difference? In the limiting case it could be that *nothing but* the difference in the proportion of altruists in groups accounts for this marked difference in fitness between identical genotypes in environments that are identical aside from the group structure. In such a case you have to look at



the patterns of interaction (S&W's minimalist definition of a group) to see why the results come out as they do.

But who is defining what into, or out of, existence? S&W use Simpson's paradox to motivate their discussion of the importance of groups. The non-biological example they use is so telling that it is worth quoting in full, for philosophers unfamiliar with it:

During the 1970s, the University of California at Berkeley was suspected of discriminating against women in its graduate admissions policies.... The percentage of women applicants who were admitted was less than the percentage of men, and the difference was large enough that it could not be attributed to chance. The University conducted a department-by-department inquiry and found in each department that women were admitted no less often than men. Evidently, women did worse than men overall, but not in any department.... The paradoxical finding fell into place when it was noticed that women tended to apply to departments with low acceptance rates. To see how this can happen, imagine that 90 women and 10 men apply to a department with a 30 percent acceptance rate. This department does not discriminate and therefore accepts 27 women and 3 men. Another department, with a 60 percent acceptance rate, receives applications from 10 women and 90 men. This department doesn't discriminate either and therefore accepts 6 women and 54 men. Consider both departments together, 100 men and 100 women applied, but only 33 women were accepted, compared with 60 [*sic*—57, in fact] men. A bias exists in the two departments combined, despite the fact that it does not exist in any single department, because the departments contribute unequally to the total number of applicants who are accepted. (pp. 24-6).

In their discussion of the application of Simpson's paradox to biological groups, they say "we need to emphasize that adding the progeny from the two groups is just statistical sleight of hand unless it can be justified biologically. If the two groups are permanently isolated from each other, natural selection will eliminate the altruists within each group, as we have already shown." (p. 25) In the Berkeley example of Simpson's paradox, the adding of the "progeny" was not, in one sense, statistical sleight of hand, because it gave the right answer: women *were* "doing worse than men overall, but not in any department." If we want to assign a number to play the role parallel to fitness in the biological case, which number should we use, the department-relative number or the university-relative number? In fact, we need to keep track of both, for there are different phenomena that need to be explained, and it is only sleight of hand if you confuse your audience (or yourself) about which you are using.

As S&W insist, evolutionary altruism—the topic of part I of their book—must not be confused with psychological altruism—the topic of part II. But it is not clear to me that they themselves keep the unwarranted psychological implications from wafting into their discussions in part I. Consider their discussion of the fanciful example introduced by a cartoon showing a school of small fish taking on the shape of a very large fish and thereby driving off a large potential predator.



... it is not obvious that altruism is required for this process to occur. That would depend on how the costs and benefits of assuming the fish shape are distributed among members of the group. Perhaps some fish *voluntarily* [emphasis added] assume the most dangerous positions in the school, in which case altruism would be involved. However, it is equally possible that all positions are equally dangerous, or the fish have some way of randomizing their position, or some fish are *forced* [emphasis added] by others to assume the most dangerous positions. Regardless of the degree of altruism involved, we can appreciate that the school is well-adapted at the group level. [p. 31]

What are the words “voluntarily” and “forced” doing in this passage? Evolutionary altruism has nothing to do with psychological states or motives—it is, after all, as accessible to daisies and slime molds as it is to minnows. Is there any way of distinguishing, at this level, between an *altruist* and a *chump*?¹ The connotations are opposite, but if we leave out motivation and knowledge, the profile is apparently the same. Sober, in correspondence, clarifies their definition of evolutionary altruism thus: “any proximate mechanism (mindless or psychologically altruistic or psychologically egoistic or psychologically chumpy, if these are in fact behaviorally equivalent) that leads organisms to donate fitness benefits at cost to self.” But even the word “donate” is loaded. Think about paying taxes: When would this count as altruistic and when as prudent (a selfish investment in security)? And when would it count as simply stupid (a bad investment in security)? As we observe the school of fish scrambling into its big-fish formation, what shall we look for to tell if those assuming the most dangerous positions have volunteered or been forced?

The absence of “force” can only be a matter of there not having been any costly move (action, behavior,...) by the other fish that has the effect of inducing or enforcing that dangerous choice. He who pays his taxes under threat of a severe beating has a short-range prudential reason for doing so. That is not volunteering. Then how about those who engage in the risky behavior of doing the threatening? Are they evolutionary altruists? Not if their neighborhood has a policy of punishing those who refrain from threatening to punish tax-evaders. And so on.

On re-reading this brief essay, I note that it is bristling with unanswered questions. Let me close by suggesting that readers should also imagine there to be tacit question marks in parentheses after the few assertions that I venture here. I am that far from thinking that I have these excellent issues figured out, and I expect my confusions will once again be diminished by S&W’s response, as it has in our many earlier exchanges, for which I am grateful.

¹ Thanks to Gabrielle Jackson for this point, and the vivid term.

