

# **WHAT A CONTRACTARIAN MUST MINIMALLY BELIEVE ABOUT PREFERENCES AND THEIR REVISION**

**Paul Vimitz**

Philosophy, U. of Lethbridge

[vimip0@uleth.ca](mailto:vimip0@uleth.ca)

for the CARP Conference at St. Mary's University in Halifax on October 17, 2004

## **Abstract**

**Having characterized contractarianism and the role the notion of preferences plays in it, I provide an ontology for preferences which does away with a number of quandaries, including**

- a) whether to privilege revealed preferences or reported ones,**
- b) whether preference revision is a one-step process or a two-step one, and**
- c) whether the notion of false consciousness is intelligible.**

# WHAT A CONTRACTARIAN MUST MINIMALLY BELIEVE ABOUT PREFERENCES AND THEIR REVISION

## I. WHAT CONTRACTARIANISM IS AND IS NOT

To begin, three points of clarification.

First, normative (a.k.a. prescriptive) contractarianism is a theory about what moral intuitions a rational agent should inculcate in herself, and/or what political institutions and/or arrangements she should promote.

The word 'should' here is being used in its prudential/hypothetical sense. I call this sense the 'prudential/hypothetical' because prudence is a relation between means and *ends*, and because - save, trivially, for the end of pursuing one's ends - all ends are hypothetical, as distinct from categorical. So since, as I say, there are no non-trivial categorical ends, there's nothing it would be prudent to do categorically, save trivially to be prudent. So contractarianism is a theory about what moral intuitions a rational agent should inculcate in herself, and/or what political institutions and/or arrangements she should promote, given her *ends*.

Second, when people have ends in common there can be no need for morality.<sup>1</sup> So the contractarian mis-speaks herself if she says we all have our own self-interest in common. For if self-interest is something we could have in common, morality couldn't emerge from self-interest. And yet contractarians hold that morality *does* emerge from self-interest. So what we should say is that I have an interest in my interests and you have an interests in your interests. But it would commit Mill's Fallacy to say we therefore have interests in common. For my interests and your interests are two very different interests.

---

<sup>1</sup> There can be at most a need to solve a coordination problem. Coordination problems are as difficult as moral ones, but it pays not to conflate the two kinds of problems.

And third, the etymology of the word notwithstanding, contractarianism has nothing to do with contracts. More specifically - and contrary to how many contractarians mis-describe their own project - contractarianism is *not* the view that the right and/or just is what rational agents would *agree* to, whether that agreement be made behind some veil of ignorance or in front of it, or whether it be hypothetical, historical, actual, or whatever.<sup>2</sup> In Orwell's *1984*, when Goldman threatened to open the rat cage, it would have been rational for Winston to have agreed to virtually *anything*. But that doesn't make what he might have agreed to right or just. Rather what's right or just is what Winston should *do*. What he should do while Goldman's hand is on the latch is whatever Goldman tells him to do, in which case Winston's agreement to do as he's told bears only an *epiphenomenal* relation to what he should do. But the moment the rats are no longer a threat there *is* no relation between what Winston agreed to when they *were* a threat and what he should do now that they're not. So either way, contractarianism has nothing to do with agreement.

What *does* it have to do with? Well, by *descriptive* contractarianism is meant the view that

**both**

**a) our most heartfelt moral intuitions and  
b) our most cherished political institutions/arrangements,  
can best be understood as  
a<sup>1</sup>) intra-mental and  
b<sup>1</sup>) extra-mental  
responses respectively to otherwise sub-optimal patterns of mixed motive interactivity.**

And so by *normative* (a.k.a. prescriptive) contractarianism is meant the view that

**the *advisability* of both**

**a) our most heartfelt moral intuitions and  
b) our most cherished political institutions/arrangements,  
can best be *evaluated* as  
a<sup>1</sup>) intra-mental and  
b<sup>1</sup>) extra-mental  
responses respectively to otherwise sub-optimal patterns of mixed motive interactivity.<sup>3</sup>**

## II. WHY PREFERENCES MATTER

---

<sup>2</sup> Many of my fellow CARPers are guilty of this mis-representation of our position. Moral suasion seems not to have done the job, so death squads have been commissioned to cull the herd of these heretics. Accordingly, to them say I: Repent or do not open your door after midnight.

<sup>3</sup> The game theoretic reduction of the contractarian project is only as old as Gauthier's *Morals by Agreement*, 1986. Still, it's not a controversial reduction. That is, I know of no contractarian who'd deny the reduction. And not just because I'm defining contractarianism *by* the reduction. For example, Hobbes is a contractarian notwithstanding he wouldn't have had a clue what to make of the game theoretic reduction of his thinking. Thus if someone wished to define contractarianism in Hobbesian terms, I'd have no objections.

But mixed motivity, in turn, is defined in terms of agents' *preferences*. For example, in Prisoners' Dilemma, though our second and third preferences coincide, my first is your fourth and vice versa; and it's in virtue of this order to our *preferences* that our motivations are mixed. So the intelligibility of the contractarian schema - be it justificatory (i.e. normative) or merely explanatory (i.e. descriptive) - rests on the intelligibility of the notion of preferences.<sup>4</sup>

And, of course, on there *being* such things. That is, most of us understand what it would be to be a unicorn. We just don't think there *are* any. So one might understand what it would *be* to have preferences and yet deny anyone *has* any. Not only that, one might think that, though we both have 'em and understand what they are, our preferences aren't stable or well-behaved enough to underpin the contractarian reduction of ethics and politics to interactive rationality. So, it would seem, we stand in need not only of

- 1) an *analytics* for our notion of preferences, but also
- 2) the set of conditions these preferences must satisfy if they're to serve as the justificans/explanans of our reduction of morality and politics to interactive rationality. And, finally, we need
- 3) proof that these conditions actually obtain.

This is a tall order. Let's see if we can fill it.

### III. THEIR ONTOLOGY

First, though one *can* be a realist about preferences, a contractarian *needn't* be, provided only she's an *instrumentalist* about them. To explain:

---

<sup>4</sup> One common mistake - though certainly an understandable one - is to conflate preference with welfare. This is an understandable mistake because people whose preferences are disconnected to their welfare tend not to reproduce this disconnect. So it's not surprising that there's a pattern to such connection. But if we were to peg the rationality of our choices to our welfare rather than our preferences, we'd first need to settle on a theory of welfare, a.k.a a theory of the good. And yet contractarians, as such, have been loath to do this.

That said, some contractarians believe the good just *is* the satisfaction of preferences. Others espouse an independent theory of the good. But even the latter won't countenance the conflation of the two notions. A choice is rational just in case it's designed to maximize on the satisfaction of one's preferences, no matter *how* injurious those preferences might be to one's well-being. One can talk intelligibly about the *rationality* of one's preferences only insofar as one can talk - as we will below - about the rationality of altering one's preferences in the light of other *extant* preferences.

A preference is, at the very least, a relation between two mental states. More particularly it's a relation between two mental representations of states of affairs. More particularly still, it's

**the relation of being more disposed to take measures  
to actualize one such state of affairs than the other.**

Why do I say “more disposed to *take measures to actualize ...*” rather than just, say, “more favourably disposed to ...”? Because it's not clear what it would be to be more favourably disposed to some state of affairs beyond - or even as distinct from - being more disposed to take measures to actualize it.

For example, we talk of Abel's sacrifice being more pleasing to God than Cain's. But if all God's being more pleased by Abel's sacrifice amounted to was His *saying*, “I like your brother Abel's sacrifice more than yours!”, it's unlikely Cain would have had cause to murder his brother. Cain murdered his brother because he thought God might *do* something as a consequence of this partiality, that He might *reward* him and his brother differentially. And so, since preferences matter only insofar as they can be expected to eventuate in differential action, preference is a measure of one's disposition to *take measures to actualize* a state of affairs.<sup>5</sup>

But - and herein lies the *instrumentality* of talk about preferences - the mental representation of a state of affairs, in turn, is, whatever else it might be, a *postulate* of interpersonal understanding. That is,

**A prefers X to Y**

is at minimum shorthand for

**B's computational and time constraints are such that  
it behooves B to attribute to A the preference for X over Y<sup>6</sup>**

As such, this attribution says as much if not more about B than it does about A. For example, since, ex hypothesi at least, God's computational and time constraints are such that He need never take up the intentional stance with respect to any of His creatures - since, that is, God can analyze our behaviour at the sub-electronic level of description - you and I don't have preferences as far as God's concerned. By contrast, if Socrates were transported in a time machine to the present, he'd be more than entitled to think of the thermostat that it had a preference to keep the temperature of the

---

<sup>5</sup> Even when my father was on his death bed I wanted him to be proud of my latest accomplishments. So I don't mean to suggest that just because someone might be unable to do anything to us we don't care what he or she thinks of us. But I *do* mean to suggest that my concern for what he thought of me under those circumstances wasn't a *moral* concern - though it's probably true that as a matter of *psychological* history I cared what he thought of me in *part* because I'd been *habituated* to caring what he thought of me from the time when he *could* do things to me. But I also suspect I cared because I was concerned *intrinsically* about his state of contentedness. And it's *that* which I'm claiming isn't a moral concern.

<sup>6</sup> Note that we don't preclude the possibility that A=B.

room constant.

So preferences are attributions. Whether in addition they're *real* - in whatever sense of 'real' the realist about mental states has in mind when she insists on the reality of mental states - qua contractarians, at least, we need neither know nor care.

#### IV. REVEALED VERSUS REPORTED

What we *do* need to care about, however, is *how* we go about attributing the preferences we attribute, both to our co-players and, of course, to ourselves<sup>7</sup>. I *say* I prefer Coke to Root Beer, and perhaps I even believe that I do. But suppose every time a Coke and a Root Beer are placed before me I reach for the Root Beer rather than the Coke. What should you conclude about which I prefer? For that matter, what should *I* conclude about which I prefer?

Another way to put the issue is to ask whether one can or cannot act contrary to her own preferences. If she can, and if she does, is she acting irrationally? If she does act irrationally, is she no longer a rational agent? And if she's not a rational agent, does she no longer fall within the scope of the contractarian's theory? If on the other hand we say she *can't* act contrary to her own preferences, then is *that* she can't act contrary to her own preferences an analytic observation or a synthetic one?

---

<sup>7</sup> See note # 6 above.

These are important questions. But it's not clear that anything peculiar to *contractarianism* hangs on how we answer them. Utilitarianism and Kantianism are saddled with the same difficulty. So, for now at least, I'm going to take the lead of Jan Narveson, who once observed<sup>8</sup> that the moral and/or political philosopher is entitled to assume that

**there exists a defensible theory of the mind  
that will render preference-talk intelligible**

without being required, at least *qua* moral and/or political philosopher, to articulate what that theory might be.

So, in answer to query # 1 above, a preference is

**an attribution of a predilection to take measures to  
actualize one state of affairs rather than another.**

In the absence of such attributions we'd be stymied in situations involving strategic choice. But we're *not* - or at least not typically - stymied in situations involving strategic choice. So we must be making such attributions all the time. And so - in answer to query # 3 - preferences there must be.

Moreover, the grounds for the attributions of these preferences must be such that the preferences in question are stable and well-behaved enough to justify and/or explain our having adopted these aforementioned heartfelt moral intuitions and cherished political institutions and/or arrangements. Otherwise preference-talk would never have got off the ground. Hence, in answer to query # 2, preferences do have the requisite properties. So, *quod erat demonstrandum*, there are preferences, and they do precisely the work contractarianism requires them to do.

Of course it might be suggested that, by parity of reasoning, correlations between the position of the stars and how well your date will go tonight must be strong enough to justify astrology, since otherwise astrology would never have got off the ground. Astrology is a well-established predictive tool. Hence these correlations *are* strong enough. Since this is a notoriously *bad* piece of reasoning, it's likewise bad to reason isomorphically with respect to preferences.

But this is too quick. Astrology can lay claim to *very* limited success. And what little success it *can* lay claim to is readily attributable to *other* regularities that are more plausibly doing the work. Moreover, the positions of the stars do not seem to supervene on those other regularities. Preference-talk, by contrast, has virtually unlimited success, and preference-talk *does* seem to supervene on whatever postulates its detractors might wish to replace it with. So we can be indifferent to the 'reality' of preferences in a way we *can't* be indifferent to the 'reality' of the causal efficacy of the stars on our love-lives. Preferences are *serviceable* postulates. And, as already noted, they're *determinable* postulates. I prefer Root Beer to Coke just in case *you'd* be well-advised to have plenty of the former on hand if I'm one of the several people you've invited over for the weekend and you've undertaken to provide the soft drinks.

Let's call this the Dennettian Test for preference attribution. What follows from it? What

---

<sup>8</sup> albeit in private conversation

follows from it is that, since my preference-matrix is merely a predictor for *you* - and more particularly a predictor for you of my *future* behaviour - neither my past behaviour nor my own announcements about how I'll behave in the future can be taken as *critical* of what preferences I entertain. They're both merely *evidential*. And at that neither is privileged. In fact, given this Dennettian, i.e. two-person, understanding of what a preference is, it makes no sense to wonder whether you should attend to my behaviour or my testimony as to what preference I *had* at the time of the behaviour or the testimony. Why? Because in either case it would have been too late for you to use this information to predict how I'd behave. In other words, the revealed-versus-reported debate presupposes realism about mental states. But since, as I've argued, the contractarian can be indifferent to the realism-versus-anti-realism debate, qua contractarian at least, the revealed-versus-reported debate is meaningless to her.

This indifference is a radical departure from the literature, so let me make clear my argument for this indifference. Preferences are important to contractarians only because they're predictors of agents' behaviour, including one's own. Neither past behaviour nor self-reportage are critical of how one will behave in the future. Both are merely evidential. So when the evidence conflicts, whether one should go with the behaviour or with the reportage depends, as it does in *any* case of conflicting evidence, on which of the two connections has greater inductive support. But though we can't be indifferent to which is the better grounds for prediction when they do conflict, we *can* be indifferent to any *ontological* connection between past behaviour and current behaviour, and/or any ontological connection between current self-reportage on one's expected behaviour and that expected behaviour itself. The best explanation of any correlation is that there *is* such a connection. But for all we need know or care that connection could be magic.

## V. PREFERENCE REVISION AND THE REALIST OBJECTION

That said, the essence of the contractarian's advice - be it advice about what moral intuitions we should inculcate in ourselves, or what political institutions and/or arrangements we should promote - is preference *revision*. How so? Because if the moral intuitions we currently entertain were invariably those we *should* entertain, or if the political institutions and/or arrangements currently in place were invariably those we *should* promote, then contractarianism would have no prescriptive work to do. For it to have prescriptive work to do, we must be making mistakes, and those mistakes must be amenable to correction.

But there's a problem here. If and only if preferences are regarded as (in some sense at least) real does it make sense to talk about preference revision. For if, as I've been insisting, preferences are nothing more than *attributions* of differential predilections, preference revision must amount to revisions to *attributions*. But this makes no sense. Or more accurately, though it makes sense to revise an attribution, this is not what contractarians mean by 'preference revision'. So preferences can't be mere attributions. That is, when we advise a smoker, for example, to revise her preferences, we're not telling her to change her *assessment* of her predilection to reach for a cigarette rather than a celery stick. We're telling her to change her *predilections*! We're telling her to rearrange something *inside* herself. We're telling her to rearrange whatever it is *in virtue of which* it would behoove someone to say of her that she used to prefer cigarettes to celery but now she prefers celery to cigarettes. But that aforementioned 'whatever it is' is all the realist means by her realism about

preferences. So realism about preferences cannot be superfluous to the contractarian project after all.

Call this, if you will, the Realist Objection to the Dennettian conception of preferences. How can it be parried? It can be parried by noting that the Dennettian doesn't deny that there's probably something which *accounts* for the behavioural regularities in virtue of which we say of someone that she prefers this over that. The contractarian, qua contractarian, is just indifferent to what that something might be. So when we urge upon the smoker that she revise her preferences, we're urging her to change *whatever it is in virtue of which* someone would have said of her that she preferred cigarettes to celery but would now say she prefers celery to cigarettes.

We grant that this analysis sounds odd in the context of smoking. Why not just say we're urging her to change whatever it is in virtue of which she prefers cigarettes to celery? Why bring in this second person? Why indeed, when we're talking about parametric choice?! But we want the notion of a preference to serve us as well when talking about *strategic* choice. And there the second person attribution is crucial. So to get the notion of a preference to do the work we need it to do for analyzing *strategic* interactions, we burden *parametric* choice situations with this (admitted) awkwardness.

So, the Realist Objection notwithstanding, we stick to our guns. A preference is an attribution of a differential predilection. But by a preference revision we mean a revision in the attribution only in the wake of a revision in whatever it is *in virtue of which* we make the attributions we do.

People have preferences which, by virtue of other preference they have, they ought not to have. Preferring cigarettes to celery is one such, at least for those who prefer to live longer than shorter, and so *prefer* to prefer celery to cigarettes. But the human psyche is such that we can - in fact we often do - prefer **a** to **b** notwithstanding we prefer to prefer **b** to **a**. So given that one prefers to prefer **b** to **a**, the preference for **a** over **b** is a mistake.

## VI. IS PREFERENCE REVISION ONE-STEP OR TWO?

In a Prisoners' Dilemma the psychopath prefers the categorical consequences of defecting to the categorical consequences of cooperating.<sup>9</sup> But, in the wake of the contractarian's good counsel, he *prefers* to prefer the categorical consequences of cooperating to the categorical consequences of defecting, conditional upon his co-player preferring the categorical consequences of cooperating to the categorical consequences of defecting. What should he do?

Some people think - call these people one-steppers - he should just conditionally cooperate. He should conditionally cooperate because one should do what one prefers, and if he *prefers* to prefer **b** to **a**, then, dammit, he just *does* prefer **b** to **a**!

Others - call them two-steppers - *deny* that preferring to prefer **b** to **a** just is to prefer **b** to **a**. They argue that, given that one prefers **a** to **b** and only *prefers* to prefer **b** to **a**, to conscientiously reach for **b** rather than **a** he would first have to act on the second-order desire and bring it about that he now prefers **b** to **a**. So if it should turn out that he *cannot* bring himself to prefer **b** to **a**, then he *cannot*, on pain of irrationality, reach for **b** rather than **a**.

---

<sup>9</sup> By 'categorical' here I mean regardless of what one's co-player does.

Let's see if an example will help us sort out our intuitions on this score. Consider:

I prefer Coke to Root Beer, but, because Root Beer is on sale this week, I *prefer* to prefer Root Beer to Coke. That is, if I could bring myself to prefer Root Beer to Coke I could save money *and* have my preferred soft drink. Here, it seems, preferring to prefer Root Beer to Coke is *not* the same as preferring Root Beer to Coke, any more than preferring to prefer celery to cigarettes is the same as preferring celery to cigarettes.

But wait! One could as readily argue - could she not? - that my failure to bring myself to prefer Root Beer to Coke betokens not that a preference to prefer Root Beer to Coke isn't a preference for Root Beer over Coke, but rather that I couldn't have preferred to prefer Root Beer to Coke after all, since had I *really* preferred to prefer Root Beer to Coke I would have done so. In other words, the one-step versus two-step preference revision debate presupposes precisely what the revealed versus reported debate presupposed, namely realism about preferences.

What we concluded there, recall, was that I preferred Root Beer to Coke just in case you'd be well-advised to have plenty of the former on hand if I'm one of the several people you've invited over for the weekend and you've undertaken to provide the soft drinks. What would you be well-advised to do if I only *preferred* to prefer Root Beer to Coke? If the answer is, "Nothing!", then it follows, from our Dennettian analysis of preferences, that it's simply *false* that I prefer to prefer Root Beer to Coke. So to make sense of the claim that I prefer Coke to Root Beer but prefer to prefer Root Beer to Coke, we need a case in which there's something you'd be well-advised to do in deference to my preferring Coke to Root Beer, and something both

- a) other than this and
- b) simultaneous with it

that you'd be well-advised to do in deference to my preferring to prefer Root Beer to Coke. Can such a case be concocted?

Let's repeat the question with the smoking example. Suppose I claim - as I think I do - that when I smoked I preferred cigarettes to celery but preferred to prefer celery to cigarettes, for reasons isomorphic to why I preferred to prefer Root Beer to Coke. It's clear enough what you'd be well-advised to do in deference to my preferring cigarettes to celery. But what would you be well-advised to do in deference to my preferring to prefer celery to cigarettes?

And now, for good measure, let's repeat the same question using the example of moral dispositions. The contractarian claims that though the psychopath prefers categorical defection to conditional cooperation, for reasons isomorphic to Root Beer going on sale and cigarettes being carcinogenic, he prefers to prefer conditional cooperation to categorical defection. It's clear what we'd be well-advised to do in deference to his preferring categorical defection to conditional cooperation. We should defect on him in a Prisoners' Dilemma. But what would we be well-advised to do in deference to his preferring to prefer conditional cooperation to categorical defection?

## VII. THE ANSWER

The answer, in all three cases, is, I think, the same. Or at least isomorphic. If I really prefer to prefer Root Beer to Coke, then I'll avail myself of whatever means are at my disposal to bring it about that I *do* prefer Root Beer to Coke. Suppose there are pills that will do this. Suppose you have a stash of

such pills, but these pills are very expensive. Then you'd be well-advised to remove your stash from the medicine cabinet in the bathroom if you've invited me over for dinner. Or, to complete the bi-conditional, if you'd be well-advised to have plenty of Coke on hand but also to remove your stash of soft drink preference-altering drugs from the medicine cabinet in the bathroom if you've invited me over for dinner, then it makes sense to say I prefer Coke to Root Beer but prefer to prefer Root Beer to Coke.

Similarly, then, if I really prefer to prefer celery to cigarettes, then I'll avail myself of whatever means are at my disposal to bring it about that I prefer celery to cigarettes. So though given a choice between a cigarette and a celery stick I'll reach for the cigarette, I'll also avail myself of any anti-smoking drug placed on the table before me. And so we can tell whether a smoker really wants to quit by whether he does or does not reach for that drug. If he does he does and if he doesn't he doesn't.

And so something similar can be said of the psychopath. If he prefers to prefer conditional cooperation to categorical defection he'll avail himself of whatever means are offered him to affect the requisite conversion. Some psychologists tell us psychopathy is an affective disorder that's beyond the control of the sufferer. He can do nothing about his own psychopathy, and so, since ought implies can, he's beyond the reach of the contractarian's good counsel. I have no idea whether this is true. But *if* it's true then contractarianism can at most *fine-tune* a client's morality. It cannot by suasion *make* one moral.

## VIII. PASCALIAN SELF-MODIFICATION

This procedure for fixing on preferences gets complicated, however, by Pascalian considerations. For suppose that in order to bring myself to prefer Root Beer to Coke I need to *acquire* a taste for Root Beer, and so I force myself to drink Root Beer notwithstanding I prefer Coke. Now my preferences are *underdetermined* by my behaviour. That is, I could be drinking the Root Beer because I prefer it, or I could be drinking it because I only *prefer* to prefer it and I'm drinking it to give Pascalian *effect* to that second-order preference.

What's especially interesting about Pascalian self-modification is that the moment one switches from

- 1) reaching for **b** rather than **a** only because one's trying to *develop* a preference for **b** over **a** to
- 2) reaching for **b** rather than **a** because one prefers **b** over **a**

very often goes unnoticed by observer and observed alike. For example, when I was quitting smoking I'd often *tell* myself I didn't want a cigarette when in fact I did. What it was for me to have *successfully* quit smoking, however, was for it to be *true* that I no longer wanted a cigarette. I'm pretty sure I *have* successfully quit smoking. But I have no idea when. Nor, I suspect, does anyone who observed me during the process.

But if my analysis is right, there's a test they *could* have administered, though it would have taken considerable coordination and incurred considerable expense. They could have conspired together to get me to believe that science had conclusively revealed that any connection between smoking and lung cancer had been a complete hoax. If, in the wake of being disabused of this hoax, I

reached for a cigarette, then it would have been true that I'd only *preferred* to prefer not to smoke. But if I *continued* to refrain from smoking, then I think we could conclude I really *did* prefer not to smoke.

## IX. IS 'FALSE CONSCIOUSNESS' AN OXYMORON?

What's especially interesting is cases in which one has revised a preference in light of information which she later discovers to be false. This is the plight of the REAL woman inadvertently registered for a Womens' Studies course. Conditioned from birth by patriarchy to want nothing more than to "Stand by [her] Man", she becomes convinced that had she *not* been so conditioned she would *not* want nothing more than to stand by her man. But, dammit, she *has* been so conditioned, and so, dammit, she *does* want nothing more than to stand by her man. What is she to do?

It's interesting that what one does in such situations depends on how she feels about who or what did the conditioning. And so any 'conditioner' worth his salt will try to build *into* the desire-set of his victim features that will immunize it from deconstruction, either from within *or* without. So part of what it is to be an REAL woman is to be *grateful* to those who made her so, and resentful of and resistant to anyone who'd try to make her anything else.

But if these efforts are successful, in what sense is she the victim of *false* consciousness? It's certainly not false that she wants to stand by her man. All that can be false, if anything is, is that she *should* want to stand by her man. But it's only false that she should want to stand by her man if

- a) there's something *else* she wants - something else by virtue of which she should want *not* to want to stand by her man - and
- b) she wants that something else more than she wants to stand by her man.

But if patriarchy's done it's job, these conditions won't be met. And in fact it's only because it's very hard to completely quash the desire to be a person rather than an appendage that the women's liberation movement could have got off the ground.

But *that* one was manipulated into wanting nothing more than to stand by her man is not, in and of itself, grounds to rebel against that desire. As B.F. Skinner long since observed, *every* desire we have was instilled in us by an agenda not our own. For what would an agenda being one's 'own' amount to?!

Some people think there *is* a default desire. It's the desire that whatever my desires might be - and however contingently they might have been instilled in me - they be maximally satisfied. And that in fact without such a desire, the pursuit of *any* other desire makes no sense. That is, how can I act on my desires unless I have a desire to act on my desires? But this raises the spectre of something parallel to the Lewis Carol paradox. Need I also have a desire that the desire that my desires be satisfied be satisfied? And so on?

It's clear enough that we don't have an infinite regress of desires in our heads, any more than we have an infinite regress of justificatory premises for modus ponens in our head. So how do we bootstrap ourselves to act and think? Once again, our Dennettianism comes to the rescue. We know it's useful to think of ourselves as having an algorithm in our heads that takes us from preferences to actions, including the action of altering our preferences. And we know that to think of our behaviour as algorithmic we need to think of that algorithm as being driven by axioms. But we needn't think

those axioms are *real*, any more than we need to think of infinity as real just because a serviceable mathematics requires the postulation of infinity. Preference-talk is a mode of thinking, a mode of thinking that recommends itself to us by virtue of the work it gets done for us. If preference-talk gives rise to Godelian inconsistencies, then so much the worse for consistency.

No, counters the realist, for then reality-talk is likewise a mode of thinking, which likewise recommends itself to us by the work it does, presumably maintaining the consistency of any mode of thinking in which it's embedded. So reality is real in just the way preferences are real. Reality is *instrumentally* real.

Well, say I, then Narveson would have been right to add that the moral and/or political philosopher is entitled to assume that

**there exists a defensible theory of what it is to be *real*  
that will render the reality of preferences intelligible**

without being required, at least *qua* moral and/or political philosopher, to articulate what that theory might be.

## X. CONCLUSION

So, quick recap:

A Dennettian ontology of preferences does away with a number of quandaries, and does so in one fell swoop. We need privilege neither revealed preference nor reported ones since both are merely evidential, not criterial, of what preferences a second party might postulate to assist in predicting a first party's behaviour.

Whether a particular preference revision is a one-step process or a two- is answered by whether there really are two simultaneous preferences, one first-order the other second-order. And that, in turn, is given, once again, by whether there are Dennettian grounds for *postulating* one preference or two.

And, finally, the notion of false consciousness, useful as it may well be, is a blind for something *other* than what it purports to be, since what it purports to be is oxymoronic. To say of someone that she suffers from false consciousness is really just to say that she suffers from a consciousness which is inimical to the agenda of the speaker. *I* want it to be the case that she wants not to stand by her man. For apart from what she contingently wants, there *is* no 'she' who wants something other than what she contingently wants. Which is not to say one can't conscientiously want others to want other than they do. And so want for utterly altruistic purposes. I want my son to want a good education because I think that well-educated person will be happier than that counterpart of him who is *not* so well-educated. Yes, that that person is happy is *my* agenda. But having an agenda that someone else have something good in his life just *is* what it is to have altruistic preferences.