

VIII Strength of Will: Normative Issues

MORE ON AINSLIE'S CONDITIONALS

- If I hold out now, I will hold out in the future.
- If I don't hold out now, I won't hold out in the future.

Could these give us *information* about what I am like? Problem: is it accurate information?

An extreme case: Newcomb's problem. A bizarre billionaire offers you a choice:

- (1) Box A
- (2) Box A *and* Box B

Box B contains \$1000, placed there the night before. Box A either contains \$1m, placed there the night before, or else nothing. The billionaire is a brilliant predictor of people's choices (with a 99.9% success rate). When he decided last night what to place in Box A, he contemplated whether you would choose one box, or both. If he thought that you'd greedily choose both, he placed nothing in Box A. If he thought that you would choose only one, he placed \$1m in Box A. But that, of course, was yesterday, and nothing you can do now will affect what is in the boxes. Should you choose one box or two?

Less extreme: self-signaling behavior. Hardworking Calvinists. People keeping their hands in cold water for longer if that indicates a strong heart (Quattrone and Tversky, 1984). The difficulty is that if they see themselves as self-signaling, they interfere with the very signals that they think are giving them information. But in the temptation resisting case does that matter? There isn't any *further* state that the behavior is supposed to be evidence of. If someone behaves apparently altruistically because they want to think of themselves as an altruist, then maybe they are not truly an altruist. But if someone refrains from smoking because they want to think of themselves as someone who can refrain from smoking, then they really are refraining from smoking. Compare simple self-presentation behavior.

KAVKA'S TOXIN PUZZLE

The bizarre billionaire offers you \$1m iff you will form the intention to drink a very nasty (but non-fatal) toxin. You decide to do so. Then he says that you don't actually have to drink it, only to sincerely intend on the night before. Two questions:

- (i) can you so intend?
- (ii) even if you can, is it rational to do so?

And another question:

- (iii) is it rational to drink the toxin?

The example brings out the distinction between reasons for performing an action, and reasons for intending to form the action. (Perhaps we can get them going in the other way too: you get the \$1m iff you drink the toxin without intending to.)

TWO OTHER PUZZLE CASES

The reciprocal suitcase deal (compare Broome's wolf case).

The self-torturer (note that this case, like that of Ann, seems to involve the idea not simply that desires change as time changes, but that it changes in response to other behavior).

BRATMAN'S TROUBLEMAKING PRINCIPLES

The Linking Principle: I shouldn't form an intention that I now believe I should, at the time of action, rationally revise.

Or, more precisely:

If, on the basis of deliberation, an agent rationally settles at t_1 on an intention to A at t_2 if (given that) C, and if she expects that under C at t_2 she will have rational control of whether or not she A's, then she will *not* suppose at t_1 that if C at t_2 she should, rationally, abandon her intention in favor of an intention to perform an alternative to A. (Bratman, 1998, 64)

The standard view: My ranking now should depend on the ranking that I will make at the time.

TWO PUTATIVE SOLUTIONS

- (i) *Sophistication:* accept linking principle, and standard view. Problem: you'll give into temptation
- (ii) *Strong resolution:* accept linking principle, but reject standard view, in favor of the idea that one's ranking now should not depend upon the ranking that one would give at the time of action, but instead at the time of forming the resolution. Problem: seems to make too little of our agency at the time of action. (More like a machine that is locked into the plan.)

BRATMAN'S SOLUTION

Add another condition onto what is needed for rationality: you must meet the *no regret condition*:

- (i) were you to stick with the resolution, then at plan's end you would be glad about it;
and
- (ii) were you to fail to stick with it, then at plan's end you would regret it.

Does this help with the toxin case, and the other cases?

THE CASE OF YURI.

Yuri has managed to fall in love with both Tonia and Lara. When he is with Tonia he is convinced that she is the one, and vows his undying commitment; unfortunately things are just the same when he is with Lara. Worse still, his life is so structured that he keeps spending time

with each of them. As one commitment is superseded by another, and that by another, trust is lost all round. Clearly it would be rational for Yuri to persist in his commitment to one of the women, and to restructure his life accordingly; all of them recognize that. However, the no-regret condition isn't met. We can imagine him as a naturally contented type, who will not feel regret whomever he ends up with; in which case the second clause of the condition would not be met. Or we can imagine him as a naturally discontented type, who will feel regretful either way; in which case the first clause will not be met. Or we can imagine him as ambivalent, fluctuating between regret and happiness however he ends up; in which case neither clause will be stably met.