Rationality and the State of Nature

The future of normativity will involve the state-of-nature method. Although pioneered in 1990 by Edward Craig in his book Knowledge and the State of Nature, the power of the method is only now beginning to be properly recognized. The method asks what the point of a target normative concept would be in an idealized social setting: a state of nature. In doing so, it sheds light on the concept, often in ways hidden from traditional analytical methods.

The best way to explain and show the importance of the state-of-nature method is to use it. So I present a case study, deploying the technique to resolve a puzzle about rationality and the first person.

Two compelling principles about rationality and the first person seem incompatible. On the one hand, rationality's normativity is first-personally constrained. On the other hand, rational requirements are normatively inert from the first-personal perspective. This constitutes a puzzle.

"Rationality" is used in different ways. I'm concerned with structural rationality: coherence requirements that preclude combinations of attitudes. In this sense of the term, rationality and the first person are tightly connected. Rational appraisal is tied to the appraised agent's first-personal perspective. What structural rationality requires is relative neither to facts nor even to an agent's evidence, but to her attitudes.

Here's a scenario to bring this out. Suppose Amira intends to run a marathon and believes that eating nothing but radishes is necessary for running a marathon. This belief is both mistaken and contrary to her evidence, but she stubbornly clings to it. Nonetheless, Amira despises radishes and fails to form the intention to eat nothing else. She thereby violates the requirement of instrumental rationality (IR). Amira is guilty of two distinct errors. First, she fails to respond to her evidence. But, second, given her belief (and her end), she goes wrong in not following the monomaniacal meal plan. Capturing this second error requires an attitude-relative formulation of IR.

This point generalizes. In the structural or coherence sense of rationality, its purview constrains agents given the actual attitudes that they have, no matter how misguided. This yields the following plausible principle.

Normativity

Rationality's normativity is constrained by an agent's first personal, attitude-relative point of view.

So far, so good. Normativity by itself doesn't cause problems; the trouble arises because it seems incompatible with another, equally compelling principle.

Inertia

Rationality is normatively inert from an agent's first-personal, attitude-relative point of view; there is no first-personal normative work for rational requirements.

To motivate Inertia, notice that from the agent's own point of view, what she ought rationally to do and what she believes she ought to do coincide. As Niko Kolodny has argued, structural rationality is transparent. Once an agent has taken into account the normative force of the considerations she believes are salient, there's no normative work left for rationality.

The puzzle is to explain how both Normativity and Inertia can be true. How can rationality's normativity be constrained by the first-personal perspective when there's no work for it to do from that point of view? Moreover, the puzzle needs to be resolved in a such a way that leaves room for the concept of rationality in our normative lives. It won't do to defuse the problem in a way that radically clashes with our practices of attributing rationality and irrationality.

This is where the state-of-nature method is key. My solution involves giving what Edward Craig calls a practical explication of the concept of rationality. That is, I tackle the puzzle by asking what use such a concept has in our conceptual scheme. I contend that the point of the concept of structural rationality is to allow an agent A to appraise another agent B's attitudes from B's own point of view. Thus rationality's perspectival and attitudinal focus furnishes us with an important instrument of appraisal. Such appraisal is especially useful along two dimensions. First, it facilitates normative engagement when one doesn't know-or can't assume-what one's interlocutor's reasons are. Second, charges of irrationality are uniquely poised to move those they're levelled against. It is intuitively more difficult to brush off a charge of irrationality than an accusation of failing to respond to reasons. That's in part because in the latter case, there's almost always room to quibble about what one's reasons are, or what attitude they support. But this sort of move isn't usually available when the complaint is irrationality. In such cases, an agent's attitudes are typically granted on all sides. Since attributions of (ir)rationality ascend from appraisal at the level of reasons to the level of attitudes, they gain traction where other kinds of appraisal do not.

So how is it that Normativity and Inertia can both be true? The state-of nature solution to the puzzle is that structural rationality gets normative purchase through interpersonal appraisal of first-personal attitudes. Normativity is true because rationality's normativity is attitude-relative. It depends on the attitudes agents do have, not the attitudes they should have in either a fact- or evidence-relative sense.

Inertia is true since, by governing relations between attitudes rather than between attitudes and reasons, rational requirements do not feature in first-personal deliberation. So while rationality is transparent to deliberators, rational requirements are importantly first personally constrained. Rationality thus has a key normative role in our social lives: it allows others to criticize an agent's attitudes by her own lights. But given that this role plays out interpersonally, holding an agent accountable to rational standards is a task that falls not primarily to her, but to those around her. That's not to say that agents aren't bound by rational requirements. They are. But because of Inertia, it takes a stance that is at once interpersonal and yet relative to the appraised agent's attitudes to make rationality's normativity felt.

Thus the state-of-nature method sheds light on a difficult puzzle concerning rationality and the first person.