Rational powers and inaction

Sarah K. Paul

Abstract: This discussion of Sergio Tenenbaum's excellent book, *Rational Powers in Action*, focuses on two noteworthy aspects of the big picture. First, questions are raised about Tenenbaum's methodology of giving primacy to cases in which the agent has all the requisite background knowledge, including knowledge of a means that will be sufficient for achieving her end, and no significant false beliefs. Second, the implications of Tenenbaum's views concerning the rational constraints on revising our ends are examined.

Keywords: Sergio Tenenbaum, instrumental rationality, trying.

Rational Powers in Action is a brilliant book. It is an extensive, resourceful, enjoyably-written articulation and defense of a genuinely new theory of instrumental rationality. It seeks to overthrow the tyranny of orthodox decision theory, understood as a theory of instrumental rationality, but it does so from within a profound grasp of that tradition. Further, the book takes aim at the relatively widespread view that "future-directed intentions" are attitudes governed by distinctive rational norms of non-reconsideration and persistence. Those who are inclined to continue holding these views – like myself, in the latter case – will have to contend going forward with Tenenbaum's powerful arguments against them.

In this response, I want to focus on two aspects of the big picture that I find especially interesting, at the unfortunate expense of leaving many of the central arguments untouched. First, I will discuss Tenenbaum's methodology of giving primacy to cases in which the agent has all the requisite background knowledge, including knowledge of a means that will be sufficient for achieving her end, and no significant false beliefs. Second, I will turn to the claims that the view makes about the rational constraints on revising our ends.

philinq XI, 1-2023, pp. 87-97
ISSN (print) 2281-8618-ETS doi: 10.4454/philinq.v11i1.443
Submitted: 14 November 2022
Accepted: 18 November 2022

1. *Uncertainty, error, and trying*

I'd like to start by bringing out an aspect of Tenenbaum's approach that is not fully committed to, or explicitly defended at length, but that I think goes deep into the foundations. One major aim of the theory, as Tenenbaum characterizes it, is to vindicate the idea that practical reason extends all the way to intentional action – to what is real, and nothing short of that. Much of the book is devoted to arguing against the idea that the inputs into a theory of instrumental rationality must be mental attitudes or events like preferences, desires, intentions, or choices, understood as phenomena that are metaphysically distinct and separable from action. The central thesis is that "instrumental rationality is rationality *in action*" (2020: viii). Further, Tenenbaum argues that the principles of *ETR* Derivation, *ETR* Coherence, and *ETR* Exercise are the only basic principles that govern the exercise of our instrumentally rational powers (see Tenenbaum's précis in this journal for statements of these principles).

This means that whenever an agent is legitimately required by the principles of instrumental rationality to take means to her extended ends, and to ensure that her ends are consistent with one another, there must be relevant intentional actions going on. The theory risks extensional inadequacy if there are good reasons to doubt that whenever the agent has an extended end that is a source of instrumental pressures, there is a corresponding intentional action occurring. The *ETR* addresses this worry by employing a quite broad conception of intentional action, and by emphasizing the indeterminacy that is present in nearly every end we pursue. Tenenbaum argues that most extended actions are "gappy," in the sense that they are compatible with substantial periods of inactivity (2020: 70). We can be getting in shape, for example, without actively doing anything to contribute to that end for an extensive amount of time. Indeed, on his view, intending to do something in the future is simply an instance of intentional action in which there is a gap in the beginning, unpreceded by any active part. If I now (in winter) intend to get in shape next summer, I already count as pursuing the end of getting in shape, though all the active parts of my action have yet to occur. And (luckily for us), the end of getting in shape is indeterminate in the sense that there is quite a bit of vagueness as to what counts as succeeding or exactly when I must act to bring about success. The structure of the pursuit does not require me to do much of anything at any particular moment; I simply need to do enough exercising over time to count as being sufficiently in shape, by my own lights, at some indeterminate point in time.

Further, the principles govern actions that are in progress. And goal-directed actions in progress are subject to the so-called 'imperfective paradox': one can *be doing* something that one never ends up successfully having done. I can cur-

rently be getting in shape without ever ending up in shape. These features of the intentional pursuit of indeterminate ends, characterized in the progressive, collectively serve to break down the barrier that intuitively exists between having an end and actually acting in pursuit of it.

At the same time, we might worry that this way of thinking about intentional action raises a new threat of unreality, insofar as tangible progress toward one's goal is rarely required. This suggests that our powers of instrumental rationality might often fall short of leading us to actually achieve our ends. When we are operating with false beliefs, or are uncertain about how to realize our ends, the reality of effectiveness threatens to remain largely in our minds. So we might ask: just how real is the rational meant to be, according to the *ETR*? Where does Tenenbaum's view stand on the question of whether it is necessarily a defect in one's instrumental rationality to fall short of achieving one's ends?

It seems to me that the book is ambivalent about this question. On one hand, a striking feature of Tenenbaum's approach is that for most of the book, he formulates the central *ETR* Derivation principle in terms of knowledge: he assumes that the instrumentally rational agent has knowledge of some sufficient and contributory means to her ends, and no false beliefs that will interfere with her effectiveness. This choice puts the focus on the kind of case in which the agent knows just what she needs to do in order to, say, become a profitable stand-up comedian, rather than on the case in which she is uncertain about what it will take, or in which she falsely believes that her innate talent for improvisation will suffice. The assumption does much to exclude the possibility of massive failure, since it follows that the conclusion of instrumental reasoning just is the intentional pursuit of means known to be (jointly) sufficient or contributory to success. The implication is that we only exercise our powers of instrumental rationality without defect in those cases where we know how to achieve our ends and are therefore in a position to be genuinely effective.

That said, Tenenbaum gestures in the final chapter at the possibility of giving this assumption up and reformulating the Derivation principle in terms of belief rather than knowledge. The instrumentally rational agent would then be understood as deriving means to her ends by way of beliefs that are potentially false, and thus failing to be truly effective. At the same time, Tenenbaum indicates a preference to hold onto the knowledge version, thereby understanding instrumental rationality in terms of actual effectiveness. Compare a similar claim he has defended elsewhere concerning deontological theories of morality: the deontic status of an act does not depend on the agent's epistemic states (Tenenbaum 2017). When it comes to morality, we might think, we are required to keep our promises, not merely to do what we believe would amount to keeping our promises, or what would be most likely to amount to such. Likewise, the

idea would be that we are instrumentally required to take the actual means to our ends, not to do what we believe would be effective, or what would likely be effective. An agent who falsely believes there is water in his glass is failing to be instrumentally rational when he takes a drink of petrol, since this action will in fact do nothing to further his end of quenching his thirst. The power of instrumentally rational agency is the power to get things done; thus, the power is not exercised in the same way in the case of knowledge and in the case of error.

This is a fascinating conception of instrumental rationality, but also radical and in some ways counterintuitive. The thirsty agent *does* seem to be instrumentally rational in taking a drink; his practical reasoning strikes many of us as impeccable, structurally speaking, though his beliefs happen to be inaccurate. Is the *ETR* committed to this "factive" view about instrumental rationality? Tenenbaum claims not, stating that we could simply revise the minor premise of the Derivation principle to refer to the agent's beliefs rather than her knowledge. However, I want to suggest that such a revision would not in fact sit easily with other aspects of the view. To deal with the problem of false beliefs in this way would be at least potentially at odds with the way the *ETR* approaches the problem of uncertainty.

We lack knowledge of the minor premise of the Derivation Principle not only when we have false beliefs, but also when we are uncertain about how to achieve our ends. This is a relatively common situation to be in, especially with respect to high-level ends that are difficult to achieve – competitive careers, advanced degrees, long-term relationships, health, wealth, and happiness, among others. We strive to achieve these ends, but we often do not know of any means that it will suffice. And in response to such uncertainty, we sometimes formulate our intentions as disjunctive or conditional on whether some currently unknown circumstance will obtain, committing ourselves only to keeping certain options open until we figure out more specifically what we want to do. We intend things like "to pursue a PhD if we are admitted to a good program with full funding," and if not, "to either enroll in law school or go backpacking in Europe."

To address this challenge to the *ETR*, Tenenbaum denies that we *can* pursue ends if we are uncertain about how to achieve them. Rather, he argues, risk and uncertainty change the nature of the actions available to us. "If I realize that none of the means available to me can ensure that I earn a million dollars," he writes, then 'becoming a millionaire' is not a possible intentional action for me" (2020: 205). Rather, one must adopt the related end of 'trying to become a millionaire', which is a different action that involves distinct sufficient and contributory means. This resourceful move allows Tenenbaum to keep the basic structure of the view in place, since an agent who lacks knowledge of a sufficient means of E-ing may yet have knowledge of a means that is sufficient for trying to E.

However, what are the grounds for thinking that uncertainty about whether we can E prevents us from even having that end? Can't I have the end of writing a successful book even if I am uncertain about whether I can do it? (Of course, I know in some sense what it is one does in order to write a book, but I am very uncertain about whether a successful book will result if *I* take those means). The obvious thing to say in defense of this claim is that intentionally E-ing requires "practical knowledge" that one is E-ing – a claim often attributed to G.E.M. Anscombe. If one does not know how to write a good book, it follows that one could not have practical knowledge of writing it, and therefore that one could not be writing a good book intentionally. But Tenenbaum attempts to stay neutral about this Anscombean idea for the purposes of his book. And more importantly, endorsing that idea would be in tension with the possibility of revising the *ETR* to allow for instrumental reasoning to proceed by way of false beliefs. After all, the agent acting in light of false beliefs would presumably lack practical knowledge as well, at least under some descriptions that are essential for understanding what is rational about her action. The agent drinking petrol does not know he is quenching his thirst (because he isn't), and so he could not be manifesting his instrumentally rational powers in pursuit of that end.

Perhaps there is an independent motivation for the idea that trying to E is a substantively different action from doing E, one that makes no appeal to controversial claims about practical knowledge. It is true that we often talk this way (though it is not clear that Tenenbaum would want to say that we should be guided by common parlance in every case, as I'll explain in a moment). But talk can be superficial, and the important question is whether 'trying' really has an internal structure that will yield plausible results about what is instrumentally required of an agent who is trying. Note that many of the high-level ends that play an important constraining role on the *ETR* view will presumably be cases of trying. For example, the solution to the problem of the self-torturer appeals to the end of "living a relatively pain-free life." Tenenbaum also talks about the end of living "a good and happy life," understood as the joint realization of the totality of our other ends. These kinds of high-level ends will be implicated at almost all moments, and do important work by issuing permissions that allow us to violate our Pareto preferences. But surely most of us do not know of any means that is sufficient to prevent chronic, debilitating pain or deep and persistent unhappiness. We're simply trying to avoid these things. So it seems important to understand exactly what the theory says when it comes to trying.

Now, 'trying' is a very slippery concept. There is an anemic sense of trying in which it is enough to lift a finger, which means that an agent who is trying in this sense incurs almost no instrumental obligations. Tenenbaum sets this notion aside and focuses instead on a more substantive reading, which he glosses

as "doing my best to succeed under the circumstances" (2020: 214). According to the *ETR*, then, instrumental rationality in pursuit of the end of trying to E will be a matter of pursuing some means or set of means known to be sufficient for trying, understood as doing one's best under the circumstances. To understand this, we therefore need to have some grasp of what the success conditions for "doing one's best" are.

I'm not convinced that there is a determinate standard here that is internal to the structure of the activity of 'doing one's best', as opposed to the context-dependent, external standards we might use to praise or blame the agent's efforts. The agent himself will not think of his aim as 'doing his best,' or conceive of the standards of success as something other than achieving his end. Indeed, if he does not achieve his end, he will take himself to have failed in his pursuit. And he will not reason about how to do his best, under that description; this sounds like what you should do if you are trying to *appear* to have done your best, to escape censure. Rather, a rational agent who is really trying to accomplish the end will take whatever acceptable means are available to achieve the end, not merely those that will suffice for having done his best. And he will rule out any other pursuits that would cause him to fail at the end he is trying to achieve, not merely those that would cause him to fail to try. Staying out all night at a party with friends is not obviously incompatible with *trying* to complete a marathon the next day, but an instrumentally rational agent will rule this out as being incompatible (let's suppose) with *succeeding* at running the marathon.

The point is that the standards a rational agent holds himself to when he is really trying seem to derive from the end itself, and not some lesser measure of success. This makes it difficult to see why we should suppose that uncertainty necessarily renders the pursuit of that end unavailable to the agent. To be sure, the more anemic sense of trying does seem to have a different internal structure and generate few if any instrumental requirements. But the existence of the other, more committal form of trying is enough to cast doubt on the strategy of handling cases of uncertainty in the way Tenenbaum does.

We might try falling back on the idea that ordinary language encourages us to describe our actions in terms of trying when we are uncertain of success. But this would put the ETR in a difficult position with respect to other pursuits that do not fit well with ordinary language. Consider the sorts of logically complex intentions mentioned earlier, with a disjunctive or conditional structure: intending to do X if C, or to do either X or Y depending on how certain future events unfold. Such commitments are undoubtedly subject to demands of instrumental rationality; at the least, we are irrational if we do not act so as to preserve the possibility of X-ing or Y-ing should the relevant circumstances arise. Common parlance does not support the idea that there is an ongoing action to do the

needed work, however. If my intention is 'to walk to the library if Ivy is there' or 'to walk to either the library or the store', it is quite a stretch to say that I *am now* doing those things – especially if I haven't moved from my couch because I don't know yet whether Ivy is at the library. These kinds of cases suggest that Tenenbaum should not wish to put too much weight on the surface grammar of act-descriptions.

To take stock: what I have been trying to illustrate in this section is that difficult questions arise when we consider agency in the face of uncertainty, and I worry that the book treats these difficulties too lightly. Tenenbaum wants to avoid committing to the more radical interpretation of the view, according to which our instrumentally rational powers are only fully exercised without defect when we know how to bring about our ends and are thus able to be effective. But it is not so straightforward to simply reformulate the view in terms of belief or credence rather than knowledge. If knowledge is not required in order to take means to our ends, then it is unclear why we should suppose that uncertainty changes the ends we can pursue, relegating us to trying rather than doing. There are good reasons to doubt that there is always a deep distinction here from the perspective of our instrumental obligations, and the fact that we draw this distinction in ordinary language carries little weight once we notice that the ETR will need to depart from ordinary parlance in characterizing some of our more logically complex ends. The approach of treating cases of uncertainty and error as substantively different from cases of knowledge therefore seems unmotivated, in the absence of a more explicit commitment and full-throated defense of the idea that instrumental rationality should be understood in a factive way.

2. *Virtues, vices, and patterns of end-revision*

Let me now turn to a different aspect of Tenenbaum's account. First, a brief comment on Tenenbaum's treatment of the role of future-directed intentions and policies in the framework of the *ETR*. Philosophers have generally treated policies and future-directed intentions – intentions to perform an action that will begin at a later time – as attitudes of some sort. And many have thought they are the kind of thing to which norms or principles of instrumental rationality apply. For instance, some have argued that norms of structural rationality govern the coherence and persistence of our future-directed intentions over time. Perhaps we ought not to reconsider our intentions without good reason, for example, on pain of exhibiting a form of incoherence over time that will make us vulnerable to temptation and otherwise prevent us from being effective.

These claims pose a challenge to the *ETR*. In response, Tenenbaum argues that we can understand policies and future-directed intentions as extended

actions rather than attitudes, at least with respect to their internal structure. A policy of calling one's mother once a week is not relevantly different, he argues, from intentionally pursuing the end of calling her once a week (2020: 126). And as we saw earlier, he denies that future-directed intentions are fundamentally different in kind from other instances of extended action; on his view, they are simply actions in which there is a gap in the beginning, unpreceded by any active part. If we accept these conclusions, then policies and future-directed intentions turn out to be the kind of thing – extended action – to which principles of instrumental rationality can apply. That said, Tenenbaum argues extensively against the existence of non-derivative requirements enjoining intention stability or forbidding reconsideration in any particular instance. On his view, an agent can be perfectly instrumentally rational from the extended perspective, executing their intentions and policies through their actions in the knowledge that the overall pattern will suffice, without obeying any strict requirement never to reconsider or shuffle their intentions arbitrarily. They simply have to avoid doing these things too much.

This sounds eminently reasonable. But *how* do we avoid doing such things too much? Tenenbaum likes to quote Leonard Cohen lyrics to demonstrate the possibility and appeal of having a policy of faithfulness "give or take a night or two" (2020: 133). The problem is that like the lover to whom Cohen's song "Everybody Knows" was addressed, people often end up taking a lot more than a couple of nights. Tenenbaum grants that there is a place in our theory of instrumental rationality for such things as resoluteness, constancy, and self-control, but he categorizes these as instrumental virtues rather than a matter of adhering to certain principles. I'll admit to having the kind of philosophical constitution that is frustrated by talk of powers and "dispositions of the will." These sound to me like names for certain patterns of behavior, when what I want to understand is the mechanism behind those patterns. Attempting to conform to a principle is one possible mechanism for achieving an acceptable pattern, and even if the content of the principle is unjustifiably strict, the *acceptance* of that principle by the agent might be justified by appeal to its results. Viewed this way, it might be true as Tenenbaum argues that *if* we non-accidentally end up satisfying our goals and policies, we cannot be deemed instrumentally irrational for all the reconsidering, procrastinating, self-indulging, and vacillating we did along the way. And yet the best way to ensure that we non-accidentally succeed in satisfying our goals and policies might be for us to view any such lapses as problematic. In other words, the best mechanism might be overkill.

At any rate, I want to raise a slightly different question about this part of the account. In the first part of the book, Tenenbaum defends an implication of the *ETR*, which is that there are no determinate rational restrictions on how

one should revise one's ends when they come into conflict with one another. An agent in this situation can abandon either of the conflicting ends, adopt a higher-order end of giving priority to one or the other, or simply revise each of them to be more restricted so that they no longer conflict (i.e "do enough of each"). The *ETR* does not offer guidance on which way to go, and Tenenbaum claims that this is a virtue, since theories of rationality that tell us how to choose between our ends run afoul of what he calls the Toleration Constraint: a theory of instrumental rationality should avoid putting restrictions on the contents of the given attitudes, except as necessary for meeting the standards of success of these representations as defined by the theory (2020: 20).

But some instances or patterns of end-revision *are* intuitively problematic. For instance, when it comes to adjusting one's ends toward mutual compatibility, there is a difference between legitimately prudent satisficing and throwing your standards out the window. Sometimes there really is room to do well enough at everything you're committed to, but in other cases, you ought to give up at least one of your commitments rather than doing everything poorly. The distinction here belongs at least in part to instrumental and not merely substantive rationality, I think, in that the tendency to lower your standards too far is not really a way of effectively achieving all of your ends; it is more akin to akrasia. Tenenbaum himself brings up other problematic cases of end-revision in Chapter 7, where he discusses the idea of instrumental virtue and vice. He examines a case of a self-aware coward who always adjusts his ends so that he never finds himself in a position of continuing to have an end while chickening out about the means (2020: 177). Akrasia can take this form as well; when one notices that a judgment, intention or policy conflicts with the action one is really tempted to take right now, one might simply revise the pesky judgment or intention to eliminate the conflict. Inconstancy and irresoluteness can similarly occur without leading the agent to fail to take the necessary and sufficient means to any end she maintains throughout the relevant period. Thus, one of the central points of this chapter is that these problematic patterns of end-revision need not involve the failure to comply with any instrumental principle, and need not even involve *acting* irrationally. Rather, on Tenenbaum's view, they are defects in the agent's will.

I wonder whether this claim doesn't water down the initial thesis a fair bit, and put us in danger of running afoul of the Toleration Constraint. It turns out that many instances or patterns of end-revision in the face of conflict may be criticizable on broadly instrumental grounds even if they are permitted by the principles of *ETR*. And the objects of criticism are not extended actions, which means that instrumental rationality is not only a matter of "rationality in action;" it also includes dispositions of the will. Further, the *ETR* faces a challenge in explaining why some patterns of end-revision are instrumentally problematic

if they never lead to a failure to take the means to one's ends. Intuitively, the *ETR* should want to explain the coward's pattern of behavior by attributing to him the high-level end of leading a danger-free life no matter what, leading him always to prioritize his own safety. But the Toleration Constraint advises us not to criticize him on those grounds.

Tenenbaum suggests instead that the instrumental defect lies in the fact that some dispositions of the will make some ends unavailable, no matter how good the agent represents them as being. We might think, however, that the ability to render some ends unavailable to ourselves is an instrumental *virtue*, insofar as things like cowardice, temptation, and fickleness incline us to take some ends to be good when they are not, and insofar as we can recognize about ourselves that this is so. The agent who is prone to temptation will be more effective at achieving her true ends if she can render the objects of temptation unavailable to her will at the key moments. Of course, the vicious agent renders the wrong ends unavailable to herself. So we would like some way of saying, without appealing to objective facts about which ends are legitimate, that some restrictions of the will are beneficial and some defective.

As I see it, this is a major motivation behind the idea that there is rational pressure to stick with a previous decision or conform to a policy, even if it conflicts with how one views things now. Theories of practical rationality that include norms of intention non-reconsideration or persistence are in a comparatively good position to explain how we can restrict our own wills over time without making substantive judgments about the legitimacy of any particular end. Tenenbaum critiques the way this basic thought has been developed in terms of strict principles or policies, and I think his points are well taken. But I am not yet sure how radically different his solutions are, insofar as they appeal to virtues of the will that are distinct from intentional action. Either way, it turns out that a fully instrumentally rational agent must do more than simply preserve means-end coherence and consistency somehow or other, with no constraints on how she adjusts her ends in order to do so. I should note that Tenenbaum sees his account of instrumental virtue and vice as being largely independent of the main *ETR* thesis. But it does seem to me that a theory of instrumental rationality should have something to say about why certain patterns of end-revision count as problematically inconstant, irresolute, akratic, or cowardly, and it looks as though this will require resources that go beyond the internal structure of intentional action.

Acknowledgements

I am indebted to Luca Ferrero, Matthias Haase, and Sergio Tenenbaum for conversations that have greatly helped to shape the final version of these comments.

Sarah K. Paul
Philosophy Program, New York University Abu Dhabi
skp5@nyu.edu

References

Tenenbaum, Sergio, 2020, *Rational Powers in Action: Instrumental Rationality and Extended Agency*, Oxford University Press, Oxford.

—, 2017, "Action, Deontology, and Risk: Against the Multiplicative Model," in *Ethics*, 127: 674-707.