

The extended theory of instrumental rationality and means-ends coherence

John Brunero

Abstract: In *Rational Powers in Action*, Sergio Tenenbaum sets out a new theory of instrumental rationality that departs from standard discussions of means-ends coherence in the literature on structural rationality in at least two interesting ways: it takes intentional action (as opposed to intention) to be what puts in place the relevant instrumental requirements, and it applies to both necessary and non-necessary means. I consider these two developments in more detail. On the first, I argue that Tenenbaum's theory is too narrow since there could be instrumental irrationality with respect to an intention to ϕ even if one is not yet engaged in any relevant intentional action. On the second, I argue against Tenenbaum's claim that "an agent is instrumentally irrational if she knowingly fails to pursue some sufficient means to an end she is pursuing."

Keywords: instrumental rationality, means-ends coherence, intention, intentional action, trying

In his excellent book, *Rational Powers in Action: Instrumental Rationality and Extended Agency*, Sergio Tenenbaum lays out a highly ambitious, original, and powerful theory of instrumental rationality, which he calls the "extended theory of instrumental rationality" (abbreviated "ETR").¹ The five core components of that theory are stated in Chapter 2. The first is:

ETR BASIC: The basic given attitude is intentional action, more specifically, the intentional pursuit of an end. (43)

Tenenbaum notes that any theory of instrumental rationality will specify some motivationally efficacious attitude (perhaps a desire, an intention, a preference, or something similar) as its "basic given attitude." That basic given attitude will then set a "basic standard of success" for the theory of instrumental rationality (11). For instance, if *desire* is the basic given attitude, then, roughly, an instrumentally rational agent will be one who satisfies her desires. The basic given attitude isn't *itself* up for rational assessment, at least insofar as the theory

¹ All in-text parenthetical page numbers are references to Tenenbaum (2020).

of instrumental rationality goes. But it does set the standard by which we can say that someone is successful (or unsuccessful) with regard to the exercise of their instrumental rational powers. As is clear from *ETR BASIC*, Tenenbaum takes the basic given attitude to be *intentional action*.

The second and third components of the theory are its *principles of derivation* and *principles of coherence*:

ETR DERIVATION: An instrumentally rational agent derives means from ends according to the following principles of derivation:

Principle of Instrumental Reasoning (Sufficient)

Pursuing A

Pursuing B_1 & Pursuing B_2 , ..., & Pursuing B_n is a (nontrivial) sufficient means to pursuing A

 Pursuing B_i (for any i between 1 and n) (while also pursuing B_j for every j such $1 \geq j \geq n$ and $j \neq i$)

Principle of Instrumental Reasoning (Contributory)

Pursuing A

Pursuing B_1 & Pursuing B_2 , ..., & Pursuing B_n is a contributory means to pursuing A

 Pursuing B_i (for any i between 1 and n). (44)

These are principles of *reasoning* to sufficient and contributory means. But they do have, in Tenenbaum's view, some consequences for the evaluation of an agent's rationality or irrationality:

But at the very minimum we can say the following: *an agent is instrumentally irrational if she knowingly fails to pursue some sufficient means to an end she is pursuing*. (47)

The principle of coherence prohibits one from holding ends one knows cannot be jointly realized:

(3) *ETR COHERENCE*: When an instrumentally rational agent realizes that her ends are incompatible (cannot be jointly realized), she abandons at least one of the ends from the smallest subset of her ends that cannot be jointly realized. (45)

For instance, if I realize that I cannot both swim in the race and watch the soccer match, which I know is on at the same time, I'll either give up the end of swimming in the race or the end of watching the soccer match.

If we look at *ETR DERIVATION*, we see that the "basic given attitude" of *intentional action* is both a *premise* ("Pursuing A ") in the principles of instrumental reasoning and a *conclusion* ("Pursuing B_i "). The latter feature is noted in the fourth component of the *ETR*:

(4) *ETR EXERCISE*: The exercise of instrumentally rational agency is an intentional action.

The fifth and final component simply observes that the principles of derivation and coherence in (2) and (3), and any principles that can be derived from them, “exhaust the content of the principles of instrumental rationality” (47):

(5) *ETR COMPLETE*: No other basic principles govern the exercise of our instrumentally rational powers. (45)

These are the five central tenets of the *ETR*. Tenenbaum also lists out some “auxiliary hypotheses” (47) that are important to the arguments for the theory, but we’ll focus here on the central tenets.

One noteworthy feature of theory is the way in which it departs from much of the discussion of “instrumental rationality” within the literature on structural rationality. Within that literature, there is a particular focus on a requirement of means-ends coherence, which is usually formulated along the following lines:

Means-Ends Coherence: Rationality requires that [if one intends to *X*, believes one will *X* only if one intends to *Y*, then one intends to *Y*].²

If I were to intend to swim in a race tomorrow, believe that to do so I must intend to register online, but not intend to register online, I would fail to do what rationality requires of me according to Means-Ends Coherence. The brackets indicate that the requirement is a “wide-scope” requirement in that “requires” has logical scope over a conditional.³ What Means-Ends Coherence prohibits is a certain *combination* of attitudes (broadly understood to include both the attitudes one *has* and the attitudes one *lacks*): the combination of *intending to X*, *not intending to Y*, and *believing one must intend to Y in order to X*.

Means-Ends Coherence is not the only requirement of practical rationality, and, plausibly, it’s not the only requirement of *instrumental* rationality. But it’s often presented as a standard example of a structural requirement of rationality. In just looking at this formulation of the requirement, however, we can see two ways in which Tenenbaum’s theory is different. First, whereas the requirement of Means-Ends Coherence is put in place by an *intention to X*, Tenenbaum’s theory takes *intentional action* as the basic given attitude. Second, whereas Means-Ends Coherence is concerned exclusively with means believed to be *necessary* for an end, Tenenbaum’s *ETR* extends to cover both *sufficient*

² This is the formulation I work with (but ultimately suggest would need some refinement) in Brunero (2020). For a small sample of other claims regarding the structural irrationality of means-ends incoherence, or formulations of the rational requirement prohibiting it, see Setiya (2007: 668), Bratman (2009: 413), Broome (2013: 159, 169), Kiesewetter (2017: 15, 46-47), Lord (2018: 21), and Worsnip (2021: 3).

³ On wide-scope requirements, see Broome (2013: Ch. 8).

and *contributory* means. For many readers, I suspect this is a breath of fresh air. We've finally arrived at a theory of instrumental rationality that is sufficiently *practical* in that intentional *action* is both the "input" and "output" of the principles of instrumental reasoning, as sketched in *ETR DERIVATION*. And we've departed from what might seem like a peculiar philosophical obsession with *necessary* means, at the expense of consideration other instrumental relations.

I, too, welcome these developments. But I want to consider these two features of the *ETR* in more detail. In particular, in §1, I consider whether we should accept *ETR BASIC*. My central worry about *ETR BASIC*, very roughly, is that the focus on intentional action is too narrow, such that many of the central cases of instrumental irrationality, including cases that would be prohibited by Means-Ends Coherence, wouldn't be covered by the theory. In §2, I consider whether we should accept the verdicts about irrationality that Tenenbaum extracts from *ETR DERIVATION*. While I think it's not all that complicated to say what rationality requires when it comes to means believed to be necessary (here, I think something along the lines of Means-Ends Coherence is correct), matters become more complicated when we transition to thinking about means believed to be sufficient. In particular, I think there are counterexamples to Tenenbaum's claim that "*an agent is instrumentally irrational if she knowingly fails to pursue some sufficient means to an end she is pursuing*" (47) and that Tenenbaum's ingenious attempts to circumvent those counterexamples will cause further difficulties for the theory.

1. Tenenbaum tells us at the start of the book that "instrumental rationality is, roughly, a relation between intentional actions" (2). This is reflected in *ETR DERIVATION*, which has intentional actions in the role of both premise and conclusion. One way to challenge the thesis that instrumental rationality is a relation between intentional actions is to challenge the Aristotelian Thesis—that is, the thesis that intentional action is the conclusion of practical reasoning. Opponents of the Aristotelian Thesis will deny that practical reasoning concludes in an (intentional) action, and will instead insist that it concludes in an intention or a practical belief or judgment, and they would reject *ETR DERIVATION* on this basis.⁴ But I'm going to set aside that debate here, and instead consider the role of intentional action as a "premise" in *ETR DERIVATION*,

⁴ For defenses of the Aristotelian Thesis, see Clark (2001), Tenenbaum (2007), Dancy (2014, 2018), and Fernandez (2016). My own view (which owes much to Paul 2013) is that the Aristotelian thesis is mistaken, and practical reasoning concludes in either an intention or a practical judgment (see Brunero 2021). These complicated questions have been well explored by others, and would take us too far afield, so I'll leave them aside.

and as the attitude which sets the standard of (instrumental) rational success, according to *ETR BASIC*. I'll argue that the conception of instrumental rationality as a "relation between intentional actions" is too narrow, since one can be instrumentally irrational (or rational) with respect to a future-directed intention to ϕ , even if one hasn't yet engaged in any (non-mental) intentional action with respect to ϕ -ing.

It's clear that Tenenbaum wishes to contrast his theory with those theories which take some mental state to be the "basic given attitude." He writes:

So, it's not an intention to write a book, or a preference for writing a book over not writing a book, that determines that my, say, writing Chapter 2 of the book is an exercise of my instrumentally rational powers. Rather, the basic given attitude in this case is my *writing a book* (intentionally), or my *intentional pursuit* of writing a book (or intentionally pursuing the end of writing a book). (44)

One question to raise here is whether it's possible to intend to write a book without having the "basic given attitude" specified by *ETR*—that is, without engaging in some relevant intentional action. (If it's not, the contrast Tenenbaum is drawing between *ETR* and other "mental state" theories becomes less interesting.) But it certainly *does* seem possible.⁵ Suppose I'm deliberating about whether to swim in the race tomorrow, and I decide (thereby forming an intention) to swim in the race tomorrow. I'm certainly not now *swimming in the race*. (Doing so would be grounds for disqualification, since one isn't permitted to swim in the race in advance of the starter's whistle.) But nor does it seem true that I'm engaged in the *intentional pursuit* of swimming in the race (or intentionally pursuing the end of swimming in the race). At least on a fairly natural understanding of "pursue," to pursue an end would involve, perhaps among other things, the employment of measures directed toward the realization of that end. But I haven't yet undertaken any (non-mental) actions which facilitate my end of swimming in the race. All I've done is reach a decision to swim in the meet. Once I *start* employing those measures (e.g., researching directions to the meet, packing up my swim gear), it would make sense to say that I'm engaged in an intentional *pursuit* of swimming in the race (or intentionally pursuing the end of swimming in the race.). But, for now, I'm not yet pursuing any such thing.

Additionally, Tenenbaum tells us that intentional action "is an event or process in the external world" (11). And, in a passage contrasting mental actions with bodily actions, he writes: "For the purpose of proposing and evaluating a theory of instrumental rationality, we should think of intentional actions as primarily bodily actions" (15). But it certainly seems possible for me to form an intention to ϕ —perhaps I reach a decision to ϕ after deliberation—without yet

⁵ For relevant discussion, see Davidson (1978) on "pure intending."

performing any bodily actions relevant to ϕ -ing. The “event or process in the external world” is yet to come.

If it is possible to intend to ϕ without yet engaging in the intentional pursuit of ϕ -ing, this raises a concern about Tenenbaum’s theory of instrumental rationality. Suppose I’ve formed an intention to swim in the race, but haven’t yet taken those measures which would license our saying that I’m intentionally pursuing the end of swimming in the race. Intuitively, even at this early stage, there could be instrumental irrationality. If I intended to swim, but didn’t intend to register, believing this to be necessary, I would be convicted of irrationality under Means-Ends Coherence. But if Tenenbaum’s theory gets a grip only later on—once the measures needed for an intentional pursuit are undertaken—it’s unclear how it can deliver this verdict.

There are some subtleties about time and rationality that I’m passing over here. First, we need to accommodate the phenomena of “rational delay.”⁶ The updating of attitudes is a process which takes time; it can’t be done instantaneously. And so we might want to allow a “grace period” of sorts, giving the person (who intends to swim and believes intending to register is necessary) some time to form the intention to register. (It’s doubtful we’ll be able to specify the length of the grace period with any precision; we can only say that excessive slowness is not allowed.) Second, we need to accommodate the phenomena of “rational self-trust.”⁷ It may be that there’s no irrationality in failing to intend to register if one rationally trusts that one will form the intention at some later point, before it’s too late. Such temporal subtleties will be relevant to the project of arriving at a more precise formulation of Means-Ends Coherence. But they need not concern us here. Let’s just work with an example which will allow us to set them aside. First, let’s assume that I’ve intended to swim in the race, and believed intending to register is necessary, for quite some time. Maybe others have even pointed out to me that I have these two attitudes and they’ve given me plenty of time to reflect on that fact and update my attitudes, but I haven’t yet done so. Issues of rational delay do not come into play here. Second, let’s assume that it’s obvious to all involved that a decision on registering is necessary immediately—perhaps the online registration window is about to close—and so considerations of rational self-trust won’t come into play. Since I must decide now, it’s not an option to trust myself to form the intention later on. But, importantly, neither of these assumptions will involve my taking measures to promote my swimming in the race. We can still have a case in which I intend to swim in the race tomorrow (and irrationally don’t intend to register) but I’m not yet

⁶ See Podgorski (2017).

⁷ See Setiya (2007: 668).

intentionally pursuing swimming in the race. And the worry is that Tenenbaum might not have the resources to allow that the norms of instrumentality rationality get a grip this early on.

One available reply to this worry comes out of Tenenbaum's discussion of what he calls "gappy actions." Tenenbaum observes that many actions are such that we can be in the process of performing them, while not at that very moment taking steps that facilitate or promote the performance of that action (70-76). He gives the example of baking a cake. In the course of performing this action, I may engage in several other actions:

Turning the oven on
Checking the cat
 Whipping eggs
Listening to the radio
 Measuring flour (130; see Fig. 5.1)

The italicized actions are the "gaps" in my gappy action of baking the cake, since they are neither instrumental nor constitutive means to baking the cake. But once one allows for the possibility of gappy actions, there's no reason to disallow the "gap" from appearing in the *initial stages* of the gappy action. Perhaps we should think of my intentional pursuit of swimming in the race as a gappy action with an *initial gap*, and allow that the action begins at the moment I intend to swim in the race, but the instrumental (or constitutive) means are taken later on. Tenenbaum suggests a view along these lines in Chapter 5:

As I see the need to paint the fence, I could get an early start by painting the first yard, the first foot, the first inch, or just by forming the intention to paint it in the near future. Forming the intention is just the limit case of early engagement in the pursuit of certain means to an end, not any different than engaging in a gappy action, except that the relevant gap is prior to the fully active parts of the action. (124)

So, with respect to our example, we could allow that one is engaging in the intentional pursuit of swimming in the race even *before* one takes any instrumental measures that promote or constitute swimming in the race. Let's call this the "initial gap strategy."

The initial gap strategy goes some ways toward solving our difficulty. But it doesn't seem to go far enough. Suppose that I initially intend to swim in the race, but I don't *ever* take any instrumental or constitutive means to doing so. In this case, it's hard to see how we can say that there's an initial gap, since there's no other surrounding actions to give definition to that gap—that is, there's nothing parallel to turning the oven on, whipping the eggs, and measuring the flour in the earlier example, which are the instrumental or constitutive means, and

which set the boundaries of the gaps. The “gap” seems to no longer exist, much like the donut hole that disappears after the donut is consumed. More importantly, it doesn’t seem like there’s any extended gappy action of *intentionally pursuing swimming in the race* in cases in which *no* instrumental or constitutive means are taken. But this is worrisome since such cases could very well be cases in which one is instrumentally irrational. Our central example of means-ends incoherence—in which one intends to swim in the race, believes one must intend to register, but doesn’t intend to register—could be a case in which no instrumental or constitutive means to swimming are ever undertaken. This case seems to me (and to many others writing about structural irrationality) to be a case of instrumental irrationality. But it’s not clear to me how Tenenbaum’s theory can deliver that result.

So far, I’ve argued that the *ETR* is too narrow: there are central cases of instrumental irrationality that would be prohibited by Means-Ends Coherence, but wouldn’t be prohibited by the *ETR*. These cases involve agents who have intended to do something without yet engaging in any intentional *action* or *pursuit*. However, it’s worth considering ways to extend the extended theory to cover such cases. We could revise our conception of the basic given attitude by first allowing that there could be *more than one* basic given attitude, and then state that both future-directed intentions and intentional actions count as basic given attitudes for the purposes of the theory:

ETR BASIC EXTENDED: The basic given attitudes are intentional action, more specifically, the intentional pursuit of an end, *and future-directed intentions*.

The revision would have the advantage of improving extensional adequacy, in that the theory could now in principle address those cases I’m concerned about. And it seems to be a modest revision in that it wouldn’t require too much tinkering with the other components of Tenenbaum’s view. What changes would we need to make? If the basic given attitude is supposed to specify the *premises* in the principles of reasoning, we may need to make the necessary changes to the two principles of reasoning in *ETR DERIVATION*. Additionally, Tenenbaum holds that the principle of coherence is to some extent a consequence of the principles of derivation (see p. 18), so we may also have to allow that *ETR COHERENCE* applies both to the intentional pursuit of ends and to future-directed intentions as well. But this should also be seen as a welcome development, since it’s already widely thought that there’s a rational prohibition on inconsistent future-directed intentions.⁸ In short, it seems like extending the *ETR* in this proposed way would have many benefits and few costs.

⁸ As Bratman observes, there’s a requirement that our intentions and beliefs fit into a “consistent conception of the future.” See Bratman (1981: 259).

2. As I noted earlier, Means-Ends Coherence applies only to means believed to be necessary. It would be a mistake to formulate a coherence requirement along the same lines applicable to sufficient means. Consider:

Mistaken Means-Ends Coherence: Rationality requires that [if one intends to X , and believes that Y -ing is sufficient for X -ing, then one intends to Y].

Suppose I intend to donate money to some particular charitable organization, and I know there are two sufficient means to making the donation: mailing a check, and depositing an envelope with cash in the donation box. Suppose I intend to mail a check, and I don't intend to deposit the envelope. There's no irrationality here whatsoever. Yet I would be in violation of Mistaken Means-Ends Coherence: I intend to make a donation, believe depositing the envelope would suffice, but don't intend to deposit the envelope. This shows that Mistaken Means-Ends Coherence is, as its name indicates, mistaken.

Of course, this is no challenge to Tenenbaum's theory, since he doesn't endorse this view. In his view, rationality would only require, at a minimum, that one take *some* sufficient means. More precisely, his view is:

But at the very minimum we can say the following: *an agent is instrumentally irrational if she knowingly fails to pursue some sufficient means to an end she is pursuing.* (47)⁹

When I intend to make a donation, and decide upon writing a check instead of depositing the envelope, I'm still pursuing some sufficient means, and so I don't run afoul of Tenenbaum's requirement.¹⁰

⁹ One of the most interesting features of Tenenbaum's view, which I'm neglecting here since I won't have space to discuss it adequately, is (putting it very roughly) his suggestion that we move away from discussions of principles and rules of rationality to discussion of rational powers and virtues. As Keshav Singh (forthcoming) puts it, in a very insightful critical notice of Tenenbaum's book and my own, Tenenbaum offers us a "power-centric" rather than a "principle-centric" approach to instrumental rationality (whereas my own approach is, as Singh notes, firmly within the "principle-centric" tradition.) But, as Singh observes, Tenenbaum's criticism of the "principle-centric" approach involves pointing out how such principles won't deliver everything we want out of a theory of rationality, and we need to talk about rational virtues as well. But that doesn't mean that Tenenbaum rejects the enterprise of specifying principles (which is well-illustrated by his statement of a principle here, and also the two principles in *ETR DERIVATION*). And it's worth investigating whether the principle quoted here is correct.

¹⁰ One question I have about Tenenbaum's theory of instrumental rationality concerns the relationship between the principles of instrumental *reasoning* in *ETR DERIVATION* and what rationality requires according to the theory. As John Broome points out in *Rationality Through Reasoning*, an agent could engage in good reasoning, but be under no requirement to do so. (For instance, the rational requirement to believe the logical consequences of what one believes—for instance, roughly, to believe q when one believes p and $p \rightarrow q$ —applies only when one cares about the relevant question (here, the question of *whether* q). But I could very well engage in good deductive reasoning about some matter I don't care about. That would be good reasoning that is not rationally required of me.) See Broome (2013: 157-159, 247). And it seems that Tenenbaum would agree with Broome's observa-

However, there might be cases where it's rationally permissible for one to knowingly fail to pursue some sufficient means to an end one is pursuing. Consider the following case:

Principled Patty: Patty is the new chair of the Philosophy Department, and she is pursuing the end of *getting a hire*—in particular, she's aiming to get the Dean's permission to hire a logician. She knows that blackmailing the Dean would enable her to get a hire, but doing so runs afoul of her moral principles, and she refuses to do it. She instead pursues other means: lobbying members of the Dean's Hiring Advisory Committee, working on a detailed hiring request, trying to convince other departments of the value of having a first-rate logician at the university, and so forth. However, she is not sure these conventional means will be successful.

There's a difference between Principled Patty and my earlier case of the charitable donation. In the case of the donation, I know of two sufficient means to donating: writing a check and depositing the envelope. Patty, however, knows of only one sufficient means: blackmailing the Dean. The other, conventional means aren't thought by her to be sufficient, either individually or collectively, for getting a hire. It seems that Patty "knowingly fails to pursue some sufficient means to an end she is pursuing" yet she doesn't seem to be guilty of instrumental irrationality.

Now if we vary the case so that Patty thinks blackmailing the Dean is *both* sufficient and necessary, then there would be irrationality—at least if Means-Ends Coherence is correct. In that case, Patty would have the prohibited combination of intending to get a hire, believing that (intending to) blackmail the Dean is necessary, and not intending to blackmail him. But we're setting up the example such that she *doesn't* believe it's necessary, but *does* believe it's sufficient.

In Chapter 9, Tenenbaum mentions the possibility of a case structurally parallel to Principled Patty:

However, in some cases, there are no sufficient means that I know will achieve my end, but I do not abandon the end. I try means that will likely, or at least possibly, achieve my end. So, for instance, I might realize that I know of no sufficient means to achieve my

tion: after all, while there is a rational requirement corresponding to the Principle of Instrumental Reasoning (Sufficient)—the requirement to take some sufficient means—he doesn't specify any requirement corresponding to the Principle of Instrumental Reasoning (Contributory). So, he seems to acknowledge the possibility that one could engage in good instrumental reasoning according to that principle without being under any rational requirement to do so. But that raises the question of what explains why there is an associated rational requirement when there is one. Why, for instance, does the Principle of Instrumental Reasoning (Sufficient) generate a requirement to take some sufficient means, but the Principle of Instrumental Reasoning (Contributory) generate no similar requirement? It's not clear to me what the answer would be. I'll set this question aside and focus instead on Tenenbaum's view about what rationality requires when it comes to means believed to be sufficient.

end of earning a million dollars (or that the only means that I know will achieve this end, defrauding my great-uncle, is not a means I am willing to take), but that there are some actions I could perform that would have a good chance of achieving the end (becoming a lawyer) or that could at least make it possible (buying a lottery ticket). (209)

In this passage, he's primarily concerned with cases in which the agent knows of no sufficient means to achieve his end, but the suggestion in the parenthetical remark is that we could treat cases like Principled Patty (and more generally, cases in which the only sufficient means is "not a means I am willing to take") in the same way.

Tenenbaum's ingenious suggestion at this point is that in such cases, the agent's action is more accurately described as *trying to ϕ* rather than *ϕ -ing*, where *trying to ϕ* is an "essentially different action" from *ϕ -ing*. (210)¹¹ He makes the point with a different example, in which the bullies are trying to prevent the nerds from crossing the street. Tenenbaum, as one of the nerds (in the example), writes: "In such a case, it would seem that I would more naturally describe my action by saying, 'I am trying to cross the street,' rather than 'I am crossing the street.' (209-210)." And then the suggestion is that the same could be said in cases in which one is unwilling to take some sufficient means. Using our example, we could say that Principled Patty's end isn't *getting a hire*, but *trying to get a hire*—or, at least, she should revise her ends so that *trying to get a hire* is her end. As Tenenbaum puts it:

We can now say that the agent who realizes that she cannot, or is not willing, to pursue means she knows to be sufficient for her end of *ϕ -ing* must revise her ends, and among the possible acts still available to her will be the act of trying to *ϕ* . (210)

But now note that if Patty's end is *trying to get a hire*, she does indeed take some sufficient means to her end. The conventional means (lobbying the Hiring Advisory Committee, etc.) do indeed suffice for *trying to get a hire*. (They aren't sufficient for *getting* the hire, but are sufficient for trying to do so.) And thus Tenenbaum could deliver the verdict that Principled Patty is indeed instrumentally rational—she's pursuing some sufficient means to her end of trying to get a hire—thereby avoiding the objection entirely.

Tenenbaum's suggestion here is that we need to change what gets put into the "A" in the schema of Principle of Instrumental Reasoning (Sufficient), where the starting premise is "Pursuing A" and "A" is a variable for agential ends. We should have Patty's end be "trying to get a hire" and then it's easy enough to

¹¹ As he notes in a footnote on p. 209, there is a sense in which the first sentence of the previously quoted passage isn't entirely accurate: "I would now be pursuing a different action, so in some sense I did abandon the end."

maintain that Patty is indeed pursuing some sufficient means to her end, and is thus rational. I want to raise four concerns about this strategy in the remainder of this section.

My first concern is that this seems to distort Patty's practical reasoning. The "trying" is now presented as the *object* of Patty's pursuit, since we now have "Pursuing *trying to get a hire*" as the first premise in Patty's instrumental reasoning. But Patty herself would likely reject that characterization of her practical reasoning. She would likely say that *what* she is pursuing is the end of *getting a hire*, not a trying. Her trying is something that occurs *while* she is intentionally pursuing the end of getting a hire; it's not the *object* of that pursuit. The object, as she sees it, is getting a hire. Patty also knows, like the rest of us, that we aren't always successful in our pursuits.

Here's another way to think about this concern. In aiming to articulate her practical reasoning, Patty certainly wouldn't have the first premise of her reasoning be "I am pursuing the pursuit of a hire" or "I am trying to try for a hire." Such premises involve confusing redundancies, and it's not at all clear what these sentences mean. It would be much more natural for her to simply say "I'm pursuing getting a hire" or "I'm trying to get a hire." But I'm not sure that "I am pursuing trying to get a hire" is all that much better. (Just as it seems odd to say that *what* is being pursued is a pursuit, and *what* is being tried is a trying, it seems odd, though perhaps not to the same degree, to say that *what* is being pursued is a trying.) It would be much more straightforward to have "I am pursuing getting a hire" as the first premise in her reasoning, while acknowledging that this pursuit *also* involves Patty's trying to get a hire and that she knows she may or may not succeed in doing what she is trying to do.

My second concern is about how redescribing Patty's end as a trying would interact with *ETR COHERENCE*. According to that principle, "when an instrumentally rational agent realizes that her ends are incompatible (cannot be jointly realized), she abandons at least one of the ends from the smallest subset of her ends that cannot be jointly realized" (45). For instance, when I realize that I cannot both finish this paper tonight and prepare adequately for tomorrow's class, I will, if I'm instrumentally rational, give up at least one of the two ends. But I might realize these two *ends* cannot be jointly realized without thinking that the associated *tryings* cannot be jointly realized. After all, in this example, I know full well that I could give both ends my best shot and fail spectacularly at one or perhaps even both. In light of this point, the general concern is that when we redescribe ϕ -ings as tryings, we'll render *ETR COHERENCE* inapplicable to cases in which it should be applicable.

Let's apply this point to Patty's case in particular. In Patty's Department, the chair is automatically on the hiring committee, as of the very moment the hire

is approved. While Patty knows full well that she can't both *get a hire* and *not be on a hiring committee*—and so *ETR COHERENCE* would prohibit her from pursuing both ends—she doesn't believe (because it's not true) that she can't both *try to get a hire* and *not be on a hiring committee*. (These ends *are* jointly realizable, and she knows it.) And so we would need some other explanation of why she's rationally prohibited from also intending to avoid being on a hiring committee. *ETR COHERENCE* would no longer be able to deliver this result.

My third concern is more of a dialectal one. In order for this strategy to get around the original objection, it has to be the case that Patty is pursuing the end of trying to get a hire and *not also* pursuing the end of getting a hire. It's not enough to note that there's *some* description of Patty's end (the one involving trying) that has it come out that she's taking sufficient means to her end. The original problem was that there's another description of Patty's end (the one involving intentional action) that has it come out that she's failing to take some sufficient means, and the *ETR* would then declare Patty to be instrumentally irrational. To avoid that, we have to *disallow* "getting a hire" as a correct description of what Patty is doing. But this seems to be a tall order. Let's suppose that Patty succeeds in getting a hire. A third-person observer (perhaps Patty herself at a later time) might reasonably describe the instrumental means Patty undertook (lobbying the Hiring Advisory Committee, writing the detailed hiring requests, etc.) as components of the extended action of getting a hire, much like one might reasonably describe, in Tenenbaum's example, the instrumental means he took (turning the oven on, mixing the eggs, etc.) as components of extended action of baking a cake. Of course, such an observer might very well also mention a trying, but they likely wouldn't do so *at the expense of* describing the extended action; they would likely say that Patty was *both* trying to get a hire and succeeding—that is, *getting a hire*. But, as we noted above, we have to disallow "getting a hire" as a correct description. That seems to be a significant cost.

My fourth concern is a normative one. Tenenbaum thinks that the agent who is "*not willing* to pursue means she knows to be sufficient for her end of ϕ -ing must revise her ends, and among the possible acts still available to her will be the act of trying to ϕ " (210, emphasis added). This helps with Principled Patty, since we can then say that in taking the conventional means (lobbying the Hiring Advisory Committee, etc.) she is indeed taking sufficient means to her end of *trying to get a hire*, and so is rational. It gets Patty off the hook as far as the charge of irrationality goes. But do we want to allow that a *mere unwillingness* to pursue means known to be sufficient can let one off the hook in this way? Consider:

Phobic Patty: Patty is the new chair of the Philosophy Department, and she is pursuing the end of *getting a hire*—in particular, she’s aiming to get the Dean’s permission to hire a logician. Matty is the new chair of the Mathematics Department, whose first (and last) proposal as chair is to give up one of his department’s faculty lines to Philosophy, so that they can hire a logician. All Patty needs to do is walk from Philosophy Hall to Mathematics Hall and pick up the paperwork. But Patty has an intense phobia of Mathematics Hall, and refuses to walk over there and get the paperwork, even though she knows this will suffice for getting a hire. She instead pursues other means: lobbying members of the Dean’s Hiring Advisory Committee, working on a detailed hiring request, trying to convince other departments of the value of having a first-rate logician at the university, and so forth. However, she is not sure these conventional means will be successful.

Whereas Principled Patty’s unwillingness is based on good moral reasons, as is Sergio’s unwillingness to defraud his great-uncle, Phobic Patty’s unwillingness is based on an irrational fear of Mathematics Hall. But since both are equally *unwilling* to take some means they know to be sufficient, and are pursuing other conventional means to getting a hire, it seems that Tenenbaum’s theory would treat the cases alike: if Principled Patty gets off the hook, Phobic Patty does as well. But that seems to be a bad result. We want it to come out that Phobic Patty is instrumentally irrational.¹²

What the pair of examples suggests is that it can’t be that an agent’s *mere unwillingness* to take some sufficient means to getting a hire lets us instead construe the relevant end as *trying to get a hire* and then see the conventional means as sufficient for the trying (thereby removing the instrumental irrationality). Rather, she must have *good reasons* for being unwilling. Principled Patty has good reasons while Phobic Patty does not. This raises a further question of what it takes to have good reasons for refusing to take some means known to be sufficient. That might be a difficult question to answer. But there’s no principled reason for thinking that a theory of instrumental rationality couldn’t provide an answer to that question. But note that in providing such an answer, the theory would not be simply applying *ETR DERIVATION* or *ETR COHERENCE*, but would be engaging in a substantive normative inquiry about reasons.¹³ In any case, my main point here is that we need to find some grounds

¹² If the phobia is not Patty’s fault, we may not want to *blame* her for her irrationality. But it’s clear that her phobia is interfering with her rationality, and, specifically, making her instrumentally irrational with respect to her end of getting a hire.

¹³ Moving in such a direction may require that we revise *ETR COMPLETE*, which takes these two principles to be the only basic principles in our theory of instrumental rationality. Or, alternatively, it could be seen as a supplement to the two principles that helps us understand how they are applied.

for letting Principled Patty off the hook that don't extend so far as to let Phobic Patty off the hook as well.

Let's sum up the argument of this section of the paper. I've focused on Tenenbaum's claim about rationality and *sufficient means*:

But at the very minimum we can say the following: *an agent is instrumentally irrational if she knowingly fails to pursue some sufficient means to an end she is pursuing.* (47)

I've argued that Principled Patty is a counterexample, since she is not instrumentally irrational in knowingly failing to pursue the known sufficient means of blackmailing the Dean. I've then considered a reply suggested by Tenenbaum's remarks in Chapter 9—namely, that Patty (if she's rational) only has the end of *trying* to get a hire and she does take some sufficient means to that end. And I've raised four concerns about this reply: (1) it distorts the first premise of Patty's instrumental reasoning in having *trying* as the *object of her pursuit*; (2) it makes it unclear how we can apply *ETR COHERENCE* with respect to the new end (the trying, as opposed to the ϕ -ing); (3) it requires that we reject as false any third-personal report which has *getting the hire* as the relevant extended action (perhaps alongside *trying to get the hire*); and, (4) it proves too much in also letting Phobic Patty, who is also unwilling to take some sufficient means, off the hook as well.

3. In this paper, I've focused on two components of Tenenbaum's *ETR* that will be exciting and interesting to those steeped in the structural rationality literature, where Means-Ends Coherence has been a standard requirement of instrumental rationality. First, whereas Means-Ends Coherence is a requirement governing *intentions*—specifically, a requirement forbidding one from intending to *X*, believing intending to *Y* is necessary for *X*-ing, and not intending to *Y*—Tenenbaum says that “instrumental rationality is, roughly, a relation between *intentional actions*” (2, emphasis added), and the principles of reasoning in *ETR DERIVATION* are formulated to reflect that (“Pursuing A,” Pursuing B_1 ,” etc.). I've here avoided discussion of the contentious question of the conclusion of practical reasoning—specifically, of whether the Aristotelian Thesis is correct—and focused instead on the “premises” or inputs—specifically, on the idea that intentional actions, not intentions, put in place the requirements of instrumental rationality. I've argued that there's a cost to accepting the *ETR*, since many standard cases of instrumental irrationality, covered by Means-Ends Coherence, wouldn't be covered by the *ETR*. And I've argued that Tenenbaum's attempt, in Chapter 5, to remedy this difficulty by appealing to “gappy actions” with a gap at the start won't do enough to resolve the worry.

Second, whereas Means-Ends Coherence is concerned exclusively with means believed to be necessary, Tenenbaum's *ETR* is concerned with means believed to be sufficient. My suspicion is that Means-Ends Coherence has enjoyed a certain popularity in the rationality literature in part because it seems easier to say what rationality requires when it comes to means believed to be necessary, and matters become trickier when it comes to non-necessary means. And if the argument in the previous section is correct, that suspicion is confirmed to some extent. I've focused in particular on Tenenbaum's claim that instrumental rationality requires that one not knowingly fail to pursue some sufficient means to an end she is pursuing. I've presented a counterexample to that claim (Principled Patty) and argued that Tenenbaum's strategy for dealing with such cases, suggested by his remarks in Chapter 9, will generate further difficulties for his theory.¹⁴

John Brunero
Department of Philosophy, University of Nebraska-Lincoln
jbrunero2@unl.edu

References

- Bratman, M., 1981, "Intention and Means-End Reasoning," in *Philosophical Review* 90(2): 252-265.
- , 2009, "Intention, Practical Rationality, and Self-Governance," in *Ethics* 119(3): 411-443.
- Broome, J., 2013, *Rationality Through Reasoning*, Wiley Blackwell, Oxford.
- Brunero, J., 2020, *Instrumental Rationality: The Normativity of Means-Ends Coherence*, Oxford University Press, Oxford.
- , 2021, "The Conclusion of Practical Reasoning," in *The Journal of Ethics*, 25: 13–37.
- Clark, P., 2001, "The Act as Conclusion," in *Canadian Journal of Philosophy* 31(4): 481–505.
- Dancy, J., 2014, "From Thought to Action," in R. Shafer-Landau, ed., *Oxford Studies in Metaethics*, Vol. 9, Oxford University Press, Oxford.
- Dancy, J., 2018, *Practical Shape: A Theory of Practical Reasoning*, Oxford University Press, Oxford.
- Davidson, D., 1978, "Intending," in *Philosophy of History and Action* 11: 46-60.
- Fernandez, P., 2016, "Practical Reasoning: Where The Action Is," in *Ethics* 126 (4): 869-900.
- Kiesewetter, B., 2017, *The Normativity of Rationality*, Oxford University Press, Oxford.

¹⁴ Many thanks to the graduate students in my Fall 2021 Ethical Theory seminar for helpful discussion of Tenenbaum's book.

- Lord, E., 2018, *The Importance of Being Rational*, Oxford University Press, Oxford.
- Paul, S., 2013, "The Conclusion of Practical Reasoning: The Shadow Between Idea and Act," in *Canadian Journal of Philosophy* 43(3): 287-302.
- Podgorski, A., 2017, "Rational Delay," in *Philosophers' Imprint* 17(5): 1-19.
- Setiya, K., 2007, "Cognitivism about Instrumental Reason," in *Ethics*, 117(4): 649-673.
- Singh, K., in press, "New Work for a Theory of Instrumental Rationality," in *Analysis*. <https://doi.org/10.1093/analys/anac020>
- Tenenbaum, S., 2007, "The Conclusion of Practical Reason," in S. Tenenbaum, ed., *New Trends in Philosophy: Moral Psychology*, Rodopi, Amsterdam, 323-343.
- Tenenbaum, S., 2020, *Rational Powers in Action: Instrumental Rationality and Extended Agency*, Oxford University Press, Oxford.
- Worsnip, A., 2021, *Fitting Things Together: Coherence and the Demands of Structural Rationality*, Oxford University Press, Oxford.

