

Two Puzzles Concerning Spinoza's Conception of Belief

Justin Steinberg

Abstract: Spinoza's account of belief entails that if A has two ideas, p and q, with incompatible content, A believes that p (and not that q) if the idea of p is stronger than the idea of q. This seems to leave little space for dominant non-beliefs, or cases in which there is discord between one's beliefs and one's affective-behavioral responses. And yet Spinoza does allow for two classes of dominant non-beliefs: efficacious fictions [fictiones] and ideas that conduce to akrasia. I show how Spinoza can account for dominant non-beliefs within his model of cognition by distinguishing between the doxastic and the affective powers of ideas and by suggesting that doxastic power is best understood diachronically. While other scholars have stressed the elegance of Spinoza's account of ideas, this paper highlights the sophistication and flexibility of his account.

Spinoza's account of belief or judgment is developed in opposition to the Cartesian view.^a Descartes distinguishes between merely entertaining some idea and forming a belief or judgment on the grounds that the latter involves an act of will, while the former does not.¹ While Spinoza agrees that beliefs or judgments involve volitions, he denies that volitions are added to ideas by way of an extrinsic faculty of will; rather, he thinks that volitions are intrinsic features of ideas themselves (see E2p48 – 49).² This is what ideas *do*: they posit or affirm the existence of some thing. As affirmations of some content, all ideas are belief-like, for Spinoza.³

^aSpinoza uses the language of believing [*credere*] and judging [*judicare*] more or less interchangeably. He does not strictly define these concepts, but his discussion of ideas in E2p49dem. and E2p49s reveals what is distinctive about his conception of belief and judgment. Quotations from, and references to, the *Ethics* [E] and the *Treatise on the Emendation of the Intellect* [TdIE] are based on Benedict de Spinoza, *The Collected Works of Spinoza*. Vol. 1. Ed. and Trans. by Edwin Curley. Princeton, NJ: Princeton University Press, 1985. I adopt the following abbreviations for the *Ethics*: Numerals refer to parts; 'p' denotes proposition; 'c' denotes corollary; 'd' denotes definition; 'dem' denotes demonstration; 's' denotes scholium; and 'post.' denotes postulate (e.g. E3p59s refers to *Ethics*, part 3, proposition 59, scholium). Quotations from, and references to, the *Tractatus Theologico-Politicus* [TTP] are based on Benedict de Spinoza, *Theological-Political Treatise*. Trans. by Michael Silverthorne and Jonathan Israel. Ed. by Jonathan Israel. Cambridge: Cambridge University Press, 2007. Citations refer to the chapter, followed by page number (e.g. 20, 232 refers to chapter 20, page 232). All references to the Latin are to *Spinoza Opera*, edited by Carl Gebhardt. When I have included the Gebhardt numbers for clarity, I refer to the volume and page number (e.g., g ii.134 refers to volume II, page 134).

[Correction added on 27 February 2018, after first online publication: An abstract has been added.]

But if all ideas are belief-like, how does Spinoza distinguish beliefs proper from other belief-like states? Diane Steinberg has persuasively argued that the answer is that beliefs are ideas that are *stronger* than any other ideas with incompatible content that one might have.⁴ We may call this the dominance model of belief. Leaving aside the question of what exactly the strength of an idea consists in (I will return to this later in the paper), we may note here that the basic dynamics of mind assumed in dominance model of belief are widely accepted in the secondary literature.⁵

There is reason to worry, though, that Spinoza's account of ideas, including this dominance model of belief, is too crude to account for the full range of human cognitive phenomena. Perhaps worse still, there is reason to worry that the dominance model of belief conflicts with phenomena that Spinoza himself acknowledges.

In this paper, I examine two puzzles—or, more precisely, two variations of a single puzzle—concerning Spinoza's dominance model of belief. After motivating these puzzles by showing that Spinoza seems to be committed to the theses that generate them, I will attempt to show how Spinoza's account of belief can accommodate such apparently aberrant phenomena. While other scholars have emphasized the elegance of Spinoza's account of ideas, I aim to cast in sharper relief some of its sophistication and flexibility.

1. First Puzzle: Powerful Fictions

The dominance model of belief entails that:

- (1) If A has two ideas, p and q, with incompatible content, A believes that p (and not that q) only if the idea of p is stronger than the idea of q.

But despite his commitment to this conception of belief, Spinoza also seems to admit that:

- (2) It is sometimes the case that when A has two ideas, p and q, with incompatible content, A believes that p (and not that q), even when q is stronger than p.

Since these two theses look incompatible, we should consider the grounds for attributing each of these theses to Spinoza.

1.1. Evidence for (1)

While there is some scholarly disagreement concerning whether in fact Spinoza wished to distinguish beliefs from other ideas,⁶ I want to defend the view that he did allow for such a distinction, and that he did so in precisely the way indicated by (1).

Here is a simple *modus tollens* argument in defense of the distinction between beliefs and other ideas:

- P1 If all ideas were beliefs, then all non-veridical imaginings would be false beliefs.
 P2 Not all non-veridical imaginings are false beliefs.
 ∴ Not all ideas are beliefs.

However odd it may seem, the first premise is actually rather weak. For Spinoza, to imagine something is to affirm it as *actually present* (E2p17 and E2p17S). So, when one imagines something that is not present, one's imagining may be said to be non-veridical. If all ideas were beliefs, then simply imagining something that is not present would be to have a false belief. That is the claim of P1.

The critical premise, then, is P2, which states that, for Spinoza, not all non-veridical imaginings are false beliefs. To see Spinoza's reasons for embracing P2, we must consider the texts in which he discusses error. The first relevant text reads: '... the imaginations of the mind, considered in themselves contain no error... the mind does not err from the fact that it imagines, but only insofar as it is considered to lack an idea which excludes the existence of those things which it imagines to be present to it' (E2p17s). Spinoza sheds further light on this last point, concerning what error or false belief consists in, later in Part II of the *Ethics*, by way of an illustration:

...when we look at the sun, we imagine it as about 200 feet away from us, an error that does not consist simply in this imagining, but in the fact that while we imagine it in this way, we are ignorant of its true distance and of the cause of this imagining. For even if we later come to know that it is more than 600 diameters of the earth away from us, we nevertheless imagine it as near. For we imagine the sun so near not because we do not know its true distance, but because an affection of our body involves the essence of the sun insofar as our body is affected by the sun (E2p35s).

On this account, imagining itself does not itself constitute an error. One will persist in imagining (non-veridically) that the sun is near, even when one knows better, simply because of how one's body is affected. Errors, unlike mere imaginings, '[consist] in the privation of knowledge' (E2p35s), specifically the lack of powerful true ideas that would *exclude* the false idea (E2p49s; g ii.134). Corrected, or offset, imaginings are still non-veridical, but they are not errors, since, in virtue of being offset, they are not beliefs.⁷ This supports P2 above, as it suggests that one can have a non-veridical idea of imagination (e.g. that the sun is 200 feet away) without thereby having a false or erroneous belief.

What distinguishes beliefs from mere imaginings is simply that the latter are offset in some way, while the former are not. The clearest account of this comes in the rich scholium to E2p49, where Spinoza is responding to the (Cartesian) claim that belief must involve the faculty of will since one can 'feign' [*fingit*] a winged horse without believing that it exists (g ii.132 – 133).⁸ Spinoza claims that if a child had the idea of a winged horse *without any other idea*, 'he [would] necessarily regard the horse as present...for what is perceiving a winged horse other than affirming wings of a horse?' (E2p49s; g ii.134). So, if we do not believe in the existence of a winged horse when we imagine it, it is because there is in our mind some other idea(s) that 'exclude' [*tollere*] or otherwise offsets the idea of the winged horse:

[I]f the mind perceived nothing else except the winged horse, it would regard it as present to itself, and would not have any cause of doubting its existence...unless either the imagination of the winged horse were joined to an

idea which excluded [*tollit*] the existence of the same horse, or the mind perceived that its idea of a winged horse was inadequate. And then either it will necessarily deny the horse's existence, or it will necessarily doubt it (E2p49s).⁹

As Diane Steinberg has pointed out, the notion of exclusion here cannot be merely logical, since ideas with incompatible content logically exclude each other¹⁰ and yet the psychological exclusion described here can be, and perhaps must be, asymmetric.¹¹ Her plausible suggestion is that one idea (asymmetrically) excludes another when they have incompatible content and the former has greater causal power. This model of ideas as dynamic modes of thought vying for dominance in the mind not only fits with Spinoza's account of why not all imaginings are false beliefs, it coheres with Spinoza's metaphysics of ideas in general.¹² It also lends support for (1), even while it leaves unanswered exactly how the power of an idea is supposed to be understood. I will return to this issue momentarily. For now, it is sufficient to have shown that there is relatively compelling evidence that Spinoza accepted (1). What about (2)?

1.2. Evidence for (2)

Before turning to Spinoza's own views, I want to motivate the independent plausibility of (2) by considering briefly the familiar experience of going to the movies (or, if one prefers a more 17th-century-friendly experience, going to the theater). It seems correct to say that even when one is most drawn in by a film, one does not actually *believe* that the events on screen are happening. Rather, one believes that one is sitting in a theater watching a work of fiction—this is one's dominant idea, in Spinoza's sense. Nevertheless, one's affective state while watching the film may be more consonant with one's make-belief than with one's genuine belief. At least with respect to one's affective response, it seems that one's belief is actually non-dominant.

In a cluster of recent articles that draw on contemporary psychological research Tamar Szabó Gendler has limned a range of non-beliefs (e.g. forms of pretense and what she calls 'aliefs') that exhibit belief-like qualities, like affect-elicitation.¹³ Cases of belief-affect and belief-behavior discordance include: people who believe that a transparent walkway over the Grand Canyon is safe, but who (despite their apparent desire to take in the spectacle) are loath to step out onto it; people who apparently want soup and who believe that a bedpan is brand-new (direct from factory), but who refuse to eat soup from such a receptacle; and, of course, moviegoers who are affected by the films.¹⁴

Instances of belief-affect or belief-behavior discordance, which are common enough, provide *prima facie* support for the independent plausibility of (2), since the mere fact that one's affective or behavioral response is more consonant with one's non-beliefs than with one's (countervailing) beliefs suggests that, *at least in some respect*, the non-belief is actually dominant. And, it would seem, any adequate account of mind ought to be able to allow conceptual space for possibility of such dominant non-beliefs.

We find good news and bad news when we turn to consider Spinoza's own account. The good news is that he does seem to allow for dominant non-beliefs. The bad news is that it is not immediately clear how this admission can be reconciled with his theory of belief. Let's focus on the good news for the moment. Spinoza seems to

allow for two classes of dominant non-beliefs, which generate the two puzzles of this paper: (1) the class of what we might call efficacious or powerful fictions and (2) ideas that conduce to *akratic* actions. In this part of the paper we will focus on the former.

Fictions [*fictiones*] are feigned ideas. To feign [*fingere*] something to (mentally) posit [*supponere*] it without actually believing it. Spinoza's earliest, and most comprehensive, analysis of fictions is found in the middle sections of his unfinished *Treatise on the Emendation of the Intellect*. In these sections, Spinoza sets out to distinguish various ideas of the imagination—false, fictitious, and dubious ideas (*TdIE* §50)—from the true ideas of the intellect. The analysis of fictions is rather wide-ranging and at times obscure. However, certain things are clear. Fictions, as ideas of the imagination, are necessarily confused.¹⁵ But they are distinguished from false ideas in that fictions do not 'suppose assent' [*supponat assensum*] (*TdIE* §66). Spinoza claims, somewhat elliptically, that false ideas are not accompanied by causes that would prevent us from positing the extrinsic thing, whereas fictions do include such causes. Since only ideas can limit ideas, these buffering causes that preclude assent in the case of fictions must just be other ideas with sufficient offsetting power.

This fits with much of what Spinoza says about fictions in the *Ethics*.¹⁶ Consider once again E2p49s. As we saw above, he claims here that what separates a *merely* feigned idea from a belief is that the former is 'joined to' ideas that either positively exclude the existence of the feigned idea or at least render it dubious by neutralizing its power. *Mere* fictions are either positively disbelieved or doubted. I want to focus on feigned ideas of the first kind—positively disbelieved fictions—since if these can be shown to be stronger than the excluding ideas that accompany them, this will lend clear support to (2).

Spinoza writes that 'if the mind, while it imagined nonexistent things as present to it, at the same time knew that those things did not exist, it would, of course, attribute this power of imagining to a virtue of its nature, not to a vice' (E2p17s). Later we learn that this is not a mere hypothetical. Spinoza trades in fictions of this very sort near the end of the *Ethics*. He indicates that as we gain more of the third kind of knowledge, the mind becomes more eternal (E5p31s), and we begin to love God intellectually (E5p33s). He states that this view—according to which one's knowledge and love of God *begin* to be—is strictly speaking fictitious (E5p33s), since this knowledge and love of God are *themselves* eternal (E5p33dem); and eternal things, of course, do not change.¹⁷ However, Spinoza insists that this figurative manner of conceiving makes for an 'easier explanation and better understanding of the things we wish to show' (E5p31s). There is some value, then, in feigning the emergence of something that we know to be eternal.

Another apparent example of a salutary fiction can be found in the opening paragraphs of Chapter 4 of Spinoza's *Tractatus Theologico-Politicus* [TTP]. This chapter begins with an analysis of law in which Spinoza distinguishes between a law—defined as that in accordance with which things behave in fixed and determinate ways—that 'depends upon natural necessity' and one that 'depends upon a human decision' (TTP 4, 57). Given Spinoza's refusal to treat man as an *imperium in imperio*, this distinction between types of law looks strictly untenable: laws that follow from human decisions must ultimately be explicable through the laws of nature, just like anything else.¹⁸ Spinoza acknowledges this, but defends this division in part on

the grounds that it is sometimes useful to conceive of human decisions and the effects that follow therefrom as abstracted from natural necessity. He writes:

We ought to define and explain things by their proximate causes, and a general consideration of necessity and the connectedness of causes cannot help us at all in the formation and ordering of particular things. We are also ignorant of the actual coordination and connectedness of things, that is, of how things are really ordered and connected, and therefore it is better and indeed necessary for conduct of life, to regard things as possible (TTP 4, 58).

Unfortunately, Spinoza does not explain why 'it is better and indeed necessary' for us to represent things as possible rather than as necessitated. Perhaps part of the explanation seems to be that from a predictive and, especially, a deliberative perspective, the idea of an open future is an indispensable fiction. Whatever the explanation is, Spinoza encourages us to imagine things as possible rather than as necessary,¹⁹ even though he does not want this feigning to undermine one's all-out commitment to necessitarianism. Like the idea of *gaining* eternity, the idea of possibility (or the idea of an open future) is a valuable fiction, something to be imagined but not believed.²⁰

Now, whatever value such fictions have, it is obviously not on account of what they directly reveal, but rather because of some indirect good that results from how these ideas affect the mind. Fictions, like the idea of gaining eternal love of God, the idea of an open future, or the representation of moral exemplars,²¹ might stimulate us to act in ways that are more consistent with our aims than if we acted just on our beliefs.²²

Spinoza's account of mind is, in general, well suited to account for the power of pretense. If all ideas are intrinsically affirmative, or belief-like, then even fictions have a certain grip on us. What is curious here is that *disbelieved* fictions can be useful, on Spinoza's account, even while being apprehended *as* disbelieved fictions—that is, while remaining firmly offset by other more powerful excluding ideas.²³ With the more powerful idea serving as a kind of buffer, it is not clear how fictions *as* fictions could capture the mind in the way that they must in order to be salutary.

To appreciate the problem, we might contrast Spinoza's approach to fictions here with a Humean alternative, according to which, if a fiction acquires sufficient force and vivacity, it thereby constitutes a belief.²⁴ Hume states that 'the difference between *fiction* and *belief* lies in some sentiment or feeling, which is annexed to the latter, not the former' (*Enquiry Concerning Human Understanding* Section V, Part II).²⁵ Beliefs are ideas that are conceived with greater force and vivacity than fictions, which, he claims, is why histories are more vivid than novels (*A Treatise of Human Nature*, 1.3.7).²⁶ If (erstwhile) fictions were to acquire vivacity of beliefs, they would cease to be *mere* fictions. He illustrates this by noting that chronic liars can come to *believe* their lies by repeating them so frequently that they strike the mind with the same vivacity as beliefs (*Treatise*, 1.3.9, p. 81).²⁷

Spinoza's position is somewhat different, since he envisages a situation in which one is a fiction can exert sufficient power, however fleeting, over the mind to produce a salutary impact without significantly eroding one's commitment to the

countervailing belief. But this seems to run contrary to the dominance model of belief, according to which the relative power of an idea co-varies with one's level of credence in that idea. So if fictions *as* fictions gain power over the mind relative to their excluding beliefs, it would seem that, minimally, this would entail an erosion of one's commitment to these excluding beliefs. But since Spinoza indicates that we can feign things that we know [*scire*] to be false without any danger of error, it looks like his account of salutary fictions is incompatible with the dominance model of belief.

2. Partial Solution to First Puzzle

As a first step towards resolving this tension, we must scrutinize his conception of the power of ideas. The way that it is presented in much of the literature masks an ambiguity. For example, Martin Lin—borrowing an example from Michael Della Rocca—illustrates the difference between beliefs and non-beliefs by describing situation in which one has two competing ideas about the contents of a glass in front of her: one idea that it contains water and another idea that it contains turpentine. According to Lin, based on the dominance model of belief, one's belief will be revealed in one's actions (assuming that one is thirsty).²⁸ If one drinks, it is because she believes that the glass contains water; if one refrains, it is because she believes that it contains turpentine.

But, as it stands, this illustration is surely inadequate.²⁹ For, even if beliefs include a *tendency* to express themselves in action, other belief-like ideas (e.g. fictions) include this same tendency. Which of these tendencies is expressed in action will be a function of the relative strength of one's *desire*. But strength of belief does not necessarily correlate with strength of desire. One might more strongly believe that the glass contains water and not turpentine, even while being more strongly repelled by the idea of turpentine than one is enticed by the idea water. Put simply, we can distinguish between doxastic strength (strength of belief) and affective strength (strength of affect or desire) of one and the same idea.

While Spinoza does not flag up this distinction, I submit that he not only allowed for it, his account of mind depends upon it. We can see how doxastic power and affective power come apart either by considering ideas that lack affective power or by examining how the strength of an affect is determined. Let's take these in turn.

On Spinoza's view, all affects are representations³⁰; but, not all representations are affective³¹: 'There are no modes of thinking, such as love, desire, or whatever is designated by the word affects of the mind, unless there is in the same individual the idea of the thing loved, desired, and the like. But there can be an idea, even though there is no other mode of thinking' (E2a3). While affects require representation of some object; one can represent an object without being affected. Spinoza makes this point even more explicitly in *Ethics* 3 post 1: 'The human body can be affected in many ways in which its power of acting is increased or diminished, and *also in others which render its power of acting neither greater nor less*' (E3 post 1; cf. E3p15d—my emphasis). One can be affected by something, and so imagine that thing (E2p17; E2p17s), even if one's power of acting is neither increased nor diminished in this interaction, in which case one will be affectively neutral.

But even if an idea is affectively neutral, it must nevertheless possess *some* causal power (E1p36). This is sufficient to show that ideas have a power that is not coextensive with affective power. E2p49 and E2p49d suggest that this power is something like the affirmative power that is intrinsic to ideas themselves—a power to compel the mind, or doxastic power. And, as we've seen in E2p49s, Spinoza's account of doubt and disbelief depends on ideas having a certain degree of (affirmative) power in virtue of which one idea can exclude another. The entire discussion in this passage is rooted solely in the power of ideas *independent of affects*: nothing in the discussion of one's belief in a winged horse turns on one's affective attitude towards the horse. Indeed, the notion of an affect has yet to be introduced. The affirmative power of an idea which exists independently of its affective power, and which exerts influence over the mind, is just what I am calling its doxastic power.

We also get a sense of the distinction between doxastic and affective power by examining the determinants of affective power. In a couple of clusters of propositions in *Ethics* 4 and 5 (especially E4p5 – 18 and E5p5 – 9), Spinoza examines a variety of factors that determine the strength of affects. Some of these determinants of affective power described can be understood in terms of doxastic power. Take, for instance, the modal and temporal features of one's representation. Spinoza claims that affects directed towards things that are represented as necessary will be more intense [*intensior*] than affects towards things that are represented as merely possible or contingent (E4p11) because representations of things as necessary are firm and unopposed, whereas representations of things as possible or contingent include the positing of excluding causes or alternate effects. A similar analysis could be given of temporal proximity (E4p10). The steadiness or constancy of affirmation partially determines the intensity of the affect is influenced. Other determinants of affective power that Spinoza examines—including the strength of the eliciting external causes (E4p5) and, the degree of connectedness of this affect or idea to other affects or ideas (E5p8; E5p11)—while not themselves forms of doxastic power, might correlate with doxastic power; so they could be said to determine affective power just in virtue of determining doxastic power.³²

However, there is at least one determinant of affective power that is distinct from doxastic power, and that is the purity of valence. Ideas of things whose valence is unalloyed (whether negatively or positively) will be more affectively charged than ideas of things towards which we are ambivalent, since ambivalence towards an object, which Spinoza calls a 'vacillation of mind' [*fluctuatio animi*] (E3p17s), results in an attenuation, or partial neutralization, of the affective power. Such vacillations of mind are, ultimately, a different species of unsteadiness of affirmation; they are, as he puts it, 'related to the affect as doubt is to the imagination (IIP44S)' (E3p17s). The purity of affective valence plays a role in determining the intensity of an affect without itself being reducible to or dependent on doxastic power. This is an extension of the more general point, established above, that the valence of an idea is not determined its doxastic power (E3 post 1): whether I feel joy, sadness, or indifference upon representing a glass of water is *not* a function of the extent to which I affirm the existence of this water; rather, it is a function of how I represent the object of my idea as affecting my power of acting.³³

The claim, then, is not that the purity of valence is non-cognitive or does not include its own form of representation; rather, the point is that the representation of a change in one's power of acting is not reducible to doxastic power. Nothing in the features that determine doxastic power determine the valence or purity of valence of one's affective response. Because not all determinants of affective power are tied to doxastic power, it is possible for an idea to be rather doxastically weak while being affectively strong, and vice versa.

Having distinguished between doxastic and affective power, let's return to water/turpentine case. Imagine that one opts not to drink the liquid. This *could* be because the idea of turpentine is doxastically stronger; but it could also be because the idea of turpentine, though doxastically weaker, is *affectively* stronger due the purity of affective valence: one's aversion to drinking turpentine is stronger than one's desire for water, even while one more steadily affirms that the substance before one is water.

By decoupling doxastic and affective strength, we can dispel the appearance of formal inconsistency between (1) and (2), since the *strength* referred to in (1), doxastic strength, need not be the same as the strength of referred to in (2). Drawing this distinction also helps us to better appreciate what Spinoza's power-based account of belief does and does not entail, helping us to avoid crude misconstruals. Still, I think that this distinction alone does not adequately account for salutary fictions, as it would be a mistake to treat them like any other 'low credence, high affect' state. At the end of the paper, I will offer a further proposal for how Spinoza might flesh out his analysis of powerful fictions.

3. Second Puzzle: Dominant Ideas and *Akrasia*

Like the first puzzle, this puzzle concerns a class of dominant non-beliefs—in this case, ideas that lead to *akratic* actions. When we act *akratically* we act contrary to what we judge to be best at the decisive moment. Once again, this is a case where non-beliefs behave in dominant ways. And this puzzle is not amenable to the (partial) solution just offered to the first, since, as we shall see, the character of practical judgments precludes any distinction between the doxastic and the affective power of the judgment. Here are the two theses that create the puzzle:

- (3) If, at time *t*, A judges that, between two incompatible courses of action, *x* and *y*, *x* is better than *y*, then if, at that same time, A chooses (i.e. intentionally pursues) either *x* or *y*, A will choose *x*.
- (4) Incontinent actions—i.e. actions contrary to one's judgment of what is best at the moment of the decision—can occur.

3.1. Evidence for (3)

One finds what looks like a direct endorsement of (3) in Chapter 16 of the *Tractatus Theologico-Politicus*. Here Spinoza argues that agreements bind only as long as the

agreement remains in the (perceived) utility [*utilitas*] of the contracting parties. This perceived-utility account of obligation is grounded in the thesis that humans are *incapable* of abiding by an agreement that runs contrary to their perceived utility: 'it is a universal law of human nature that...of two good things every single person will choose the one which he himself judges to be the greater good, and of two bad things he will choose that which he deems to be less bad' (TTP 16, 198).³⁴ This looks like a straightforward expression of (3): we always choose what we judge to be best.³⁵

Perhaps one might wish to dismiss this endorsement of (3) as a casual and somewhat careless remark in a work that, because composed for a non-philosophical audience, did not demand rigorous formulations. But this is no mere off-handed remark. It is tied to central features of Spinoza's account of evaluative judgments and motivation. Consider, for instance, Spinoza's claim that 'we neither strive for, nor will, neither want, nor desire anything because we judge it to be good; on the contrary, we judge something to be good because we strive for it, will it, want it, and desire it' (E3p9s).³⁶ This would seem to apply, *mutatis mutandis*, to relative judgments:

One judges that *x* is better than *y* because one desires *x* more than *y*, rather than vice versa.³⁷

The claim here is that comparative judgments of goodness depend on relative strength of desire, which provides support for (3), since it supports the disputable premise (P1) in the argument below:

- P1: If A judges *x* to be better than *y*, then A desires *x* more than *y*.³⁸
 P2: If, at time *t*, A desires *x* more than *y*, then if, at that same time, A chooses (i.e. intentionally pursues) either *x* or *y*, A will choose *x*.³⁹
 ∴ If, at time *t*, A judges that between two (incompatible) courses of action, *x* and *y*, *x* is better than *y*, then if, at that same time, A chooses (i.e. intentionally pursues) either *x* or *y*, A will choose *x*.

Put baldly, choices necessarily track judgments of relative goodness because judgments of relative goodness necessarily track strength of desire. But before confidently concluding that Spinoza's account of evaluative judgments in the *Ethics* supports (3), we should pause to make sure that this is the correct interpretation of E3p9s.

All that E3p9s definitively establishes is that evaluative judgments depend *in some way* on desires, and that this dependence is asymmetric. This does not necessarily entail that evaluative judgments must *always* be rooted in (occurrent) affects or desires. Once again, it seems that there is a Humean alternative available. For Hume, of course, evaluative judgments are grounded in sentiments. But, on at least one reading of Hume, one's judgments of approbation or disapprobation need not be rooted in one's actual (occurrent) sentiments. One can correct for sentimental biases by judging from a 'general point of view,' in which case, judgments of goodness are based not on one's actual affective state, but on some idealized, or projected, affective state.⁴⁰

However, on Spinoza's account, as I read him, judgments of goodness are themselves reducible to, and explained through, affects. For convenience I will call this

the Constitution Thesis.⁴¹ According to the Constitution Thesis, what it *means* to form an evaluative judgment is just is to be affected in some particular way and to represent some thing as the cause of this affect. Evaluative judgments are nothing but affects. And the intensity of the affects like joy and sadness and the intensity of the desire co-vary.⁴²

The first indication that Spinoza embraced something like the Constitution Thesis, can be seen in E3p39s:

For we have shown above (in P9S) that we desire nothing because we judge it to be good, but on the contrary, we call it good because we desire it. Consequently, what we are averse to we call evil. So each one, from his own affect, judges, or evaluates, what is good and what is bad, what is better and what is worse, and finally, what is best and what is worst (E3p39s).

Admittedly the claim that 'each one, from his own affect [*ex suo affectu*] judges, or evaluates, what is good and what is bad' (E3p39s) on its own does not support the Constitution Thesis.⁴³ However, the subsequent illustration—'So the greedy man judges an abundance of money best, and poverty worst' (*ibid.*)—suggests that there is such a close relationship between affects and judgments that one can, as it were, read off the judgment from the affect. The covariation of judgments and affects is suggested by a later invocation of the passage: 'because each one judges from his own affect what is good and what is bad, what is better and what worse (see P39S) it follows that men can vary as much in judgment as in affect' (E3p51s). The Constitution Thesis can explain the covariation of affects and judgments as well as the priority of the former over the latter in ways that alternative interpretations cannot. However, the evidence from Part 3 of the *Ethics* is not decisive.

We find stronger support for the Constitution Thesis in 4p8, which states: 'The knowledge [*cognitio*] of good and evil is nothing but an affect of joy or sadness, insofar as we are conscious of it' (E4p8). The first thing to note here is that this is a claim about what constitutes a representation of good or evil; it is not a claim about knowledge in the sense of normative epistemology.⁴⁴ In other words, it is a claim about the nature of evaluative judgments, since to represent something as good is just to evaluate it. What Spinoza asserts here is that the representation of good or evil is *nothing but* [*nihil aliud*] an affect of joy or sadness insofar as one is conscious of it. Drawing on the reduction of 'good' and 'evil' to joy or sadness in E3p39s, Spinoza argues here —albeit dubiously⁴⁵—that a representation of good and evil is nothing other than the very affect of joy or sadness itself, and the consciousness produced therefrom.

Later applications of E4p8 support this reading. For instance, the first part of E4p19d reads: 'Knowledge [*cognitio*] of good and evil (by P8) is itself an affect of joy or sadness, insofar as we are conscious of it. And therefore (by IIP28), everyone necessarily wants what he judges to be good, and conversely, is repelled by what he judges to be evil' (E4p19d). Here, Spinoza infers that 'everyone necessarily wants what he judges to be good' from two premises. The first is the Constitution Thesis (E4p8): evaluative judgments are constituted by emotions themselves. The second, E3p28, shows that we desire (the furtherance of) whatever we regard as

conducive to joy, that is, whatever we love. The argument can thus be encapsulated as follows: Because to judge something to be good is just to love it (from the definition of 'love' and the Constitution Thesis) and because one desires whatever one loves, one will necessarily desire what one judges to be good. We see Spinoza using E4p8 in similar ways elsewhere in *Ethics* 4 (e.g. E4p15d and E4p64d). These further deployments of the Constitution Thesis reveal the extent to which his moral psychology depends on the reduction of evaluative judgments to affects. The account from *Ethics* 4 gives determinate content to the claims from *Ethics* 3 concerning the dependency of evaluative judgments on affects.⁴⁶

The Constitution Thesis can be seen as an expression of Spinoza's explanatory rationalism. As has been stressed in a good deal of recent literature, Spinoza is deeply committed to the principle of sufficient reason, abhorring as he does arbitrariness and brute facts.⁴⁷ And his account of the meaning of evaluative predicates and evaluative judgments expresses this concern. He seeks to answer the question 'what is it in virtue of which one can (truly) be said to judge that x is better than y?' For instance, what is it in virtue of which one's utterance, 'I really should go to the gym right now,' constitutes one's judgment rather than, say, idle words, or *half-beliefs*, in H.H. Price's sense.⁴⁸ Spinoza's answer is simple: one's judgment is expressed by one's dominant affective-desiderative state. In this way, Spinoza toes the judgment internalist line by insisting that one who does not desire x more than y does not *really* judge x to be better than y.⁴⁹ But whereas this can smack of stipulative or question-begging line-holding, Spinoza grounds its validity in the very metaphysics of moral judgment.⁵⁰

3.2. Evidence for (4)

Spinoza's account of *akrasia* has been the subject of a good deal of recent scholarship.⁵¹ Yet, as far as I can tell, there is no recognition within this work that Spinoza's commitment to (3) raises problems concerning the possibility of *akrasia*. Indeed, no one seems to evince any doubt as to whether Spinoza accepts thesis (4).

The clearest support for (4) comes in E4p17s, where Spinoza claims to have explained the phenomenon behind Medea's tortured declaration in Ovid's *Metamorphoses*: '*...video meliora, proboque, deteriora sequor...*' ['I see and approve the better, but I follow the worse'] (E4p17s).⁵² Still, before concluding straightaway that Spinoza allows for *akrasia* we should examine more precisely what *akratic* action consists in.

Here, it behooves us to distinguish between permissive and strict conceptions of the phenomenon. On a permissive conception we might say that one acts *akratically* when one, in some sense and to some degree, thinks that at the time of one's action it would be better to do otherwise than what one actually does. Spinoza's theory of ideas is poised to handle cases like this, since it allows that the mind may be composed of ideas with incompatible content that vie for supremacy. His main point in E4p17s and the preceding passages is that one can have a true idea of what is good and yet fail to act on it, since what determines one's behavior does not depend on the truth of one's idea; rather, it depends on the idea's affective potency (E4p7;

E4p14; E4p15). This conclusion is perfectly compatible with (3), since all that it shows is that one can have an idea that *x* is better than *y*, even while being more motivated by a contrary idea. It shows that one can *see* the better, but do the worse—which is in fact Spinoza's preferred way of describing the phenomenon.⁵³

But this permissive sense of *akrasia* is so weak as to be relatively philosophically uninteresting. Philosophers have thus focused their attention on a stricter conception of *akrasia*: cases in which one acts contrary to one's overarching, or all-things-considered, practical judgment. Now, of course, there is some ambiguity here concerning what counts as one's all-things-considered judgment. It could just mean the position that one reflectively endorses. But, in my estimation, this too is rather weak, since one's firmest judgment need not align with what one reflectively endorses. The more interesting case is one in which we understand 'all things considered' to mean the judgment that most reflects the agent's deepest doxastic commitments. The strict sense of *akrasia* consists in acting contrary to this kind of judgment. But, as far as I can tell, nothing in E4p17s or in the preceding propositions indicates that Spinoza allows for *akrasia* in this strict sense.⁵⁴

Ultimately, then, the thesis that might appear to be the most evident—thesis (4) above—could be the least plausible, if we take *akrasia* in the strict sense. But while I'm not convinced that Spinoza was particularly concerned to make conceptual space for strict *akrasia*, I do think that there is a way in which he could accept something similar to strict *akrasia* that is compatible with (3). And since this, admittedly somewhat speculative, reconciliation casts further light on how beliefs differ from other mental states, including 'fictions,' pursuing this resolution will be worth our while.

4. Partial Solution to Second Puzzle

In the solution to the first puzzle, I distinguished between doxastic and affective strength without fully examining how doxastic strength is to be measured. Doxastic strength consists in one's level of confidence in the content of one's representation. Since all ideas include the affirmation of some content, and since we will continue to affirm that content unless and until it is offset by some other idea, only offsetting ideas can erode one's confidence.⁵⁵ To the extent that one has conflicting ideas, one doubts, and doubting is tellingly described as a condition in which one 'vacillate[s]' [*fluctuat*] between the two ideas (E2p44s). So, as we considered in passing in the solution to the first puzzle, doxastic strength, or confidence, may be understood in terms of steadiness of affirmation. If, between two ideas with incompatible content, neither idea is consistently affirmed, one remains in a state of doubt or uncertainty regarding both propositions. But if one fairly steadily affirms one idea more than the other, such a firm prevalence seems to constitute a belief, even if, at any given moment, one might be more in the grip of the minority idea.⁵⁶

Let's transfer this discussion now to the realm of practical deliberation, where, as noted above, one cannot separate doxastic from affective strength. Here, the correlate to doubt is ambivalence. When one feels ambivalent towards a course of action, one undergoes a 'vacillation of mind' [*fluctuatio animi*] (E3p17s). Such

vacillations of mind are 'related to the affect as doubt is to the imagination' (E3p17s); they are another species of unsteadiness of affirmation. While we deliberate, we oscillate between competing appetites and aversions (I should keep writing. *No, I should really check my email.* I should stay in bed. *No, I should really get up and brush my teeth.* I should go to the gym. *No, I should avoid the cold*) Given Spinoza's intellectual debts to Hobbes, it is not surprising that this rather resembles the Hobbesian model of deliberation as the succession of appetites and aversions.⁵⁷

Here, as with ordinary judgments, if the intervals of saliency are more or less equivalent, it is not clear that one has a proper practical belief. However, one might have a relatively steady practical belief and yet still be moved by the contrary idea, since the decision could be made during that narrow stretch in which the minority idea is dominant.⁵⁸ We may once again think of Hobbes here: the will is the 'last appetite'; it is not necessarily the firmest, steadiest, or diachronically strongest.

The proposal, then, for reconciling (3) and (4) requires distinguishing between one's judgment about what is best as measured diachronically (in terms of the steadiness of one's preferences) and one's judgment measured synchronically. At the time of one's decision, one might represent y as superior to x , and yet one's more constant idea—that is, one's overarching belief—might still be that x is superior to y .⁵⁹ So, even if Spinoza himself was not particularly concerned to explain the possibility of strict *akrasia*, his theory of ideas is complex enough to allow for the preceding account.

5. A Concluding Note

Let me conclude by adverting to one final feature of Spinoza's theory of cognition that will unite the two parts of this paper. One of the greatest determinants of an idea's steadiness or power over time is its degree and strength of connectedness to *other* potent ideas. Spinoza indicates this in the early propositions of *Ethics* 5, for instance, when he writes: 'the more an image is joined with other images, the more often it flourishes' (E5p13).⁶⁰ If an idea is going to amount to more than a *fleeting* affirmation, it will need to be elicited or aroused with great frequency. And for this the idea will need to be tightly integrated into one's web of beliefs. Relatively isolated ideas, while they may strongly engage the mind momentarily, are not likely to exert much influence in the long term.

We can illustrate how this works vis-à-vis practical judgments with a relatively silly, but (for me) all-too-familiar, example. Suppose you intend to be a vegetarian. Typically your principle of vegetarianism and your behavior line up; that is, the practical ideal of vegetarianism has a steady affective dominance for you. However, very occasionally your commitment to vegetarianism is defeated by bacon-lust. Still, these transgressions are relatively infrequent: you don't purchase bacon, you don't find yourself daydreaming about club sandwiches, these episodes haven't led to other carnivorous acts that might further erode your commitment to vegetarianism. Bacon is not, for you at least, a gateway meat. For all its power when directly elicited (by the sight or smell of it), the desire for bacon is relatively isolated: it is not an idea that 'flourishes' because it is not well connected to other

images, affects, goals. For this reason, bacon tempts one into occasional (*akratic*) lapses, but it does not lead to the abandonment of the evaluative beliefs that oppose the consumption of bacon.

Now consider how this might work with fictions. One of the features of pretense that is often thought to distinguish it from belief is that the former takes place 'off-line.'⁶¹ Make-beliefs are cordoned off, or quarantined, from *echt*-beliefs and *echt*-desires. The metaphors of 'off-line' simulation and 'quarantining,' however limited and imperfect, can be neatly accounted for in Spinoza's cognitive scheme, and can help to explain the efficacy of some fictions. One of the quandaries in the first part of the paper concerned how fictions could exhibit the potency of beliefs without thereby becoming beliefs. Part of the explanation is that powerful fictions are quite isolated—quarantined, as it were—from one's other potent ideas. When one leaves the movie theater and the immediate stimulus for the fiction is lost, there is no network of ideas that will bolster these fictions. This is true for other fictions that one *knows* to be false: one can *feign* them, and thereby temporarily (and perhaps potentially) affirm them, without worrying that they will undermine one's (opposing) commitments, because, provided that no further compensatory adjustment are made that enable such ideas to gain a greater foothold in one's belief-system,⁶² there is a firm doxastic buffer that prevents them from exerting a steady influence. So, fictions, as isolated ideas, can be profoundly affirmed without being believed; unfortunately, so can rogue desires or representations of goodness.⁶³

Justin Steinberg
Brooklyn College, CUNY, Philosophy Department
USA
jsteinberg@brooklyn.cuny.edu

ENDNOTES

¹ See esp. *Meditation Four* (Descartes 1984). For a comparative analysis of Descartes and Spinoza on belief, see Curley 1975.

² All references to the *Ethics* are to Spinoza 1985. All references to the TTP are to Spinoza 2007.

³ The view that all ideas are 'belief-like' is widely accepted in the secondary literature. See Della Rocca 2003; Lin 2006; and Bennett 1984.

⁴ D. Steinberg 2005.

⁵ See e.g., D. Steinberg 2005; Della Rocca 2003; Lin 2006; Marshall 2013.

⁶ Della Rocca remains agnostic about whether or not all belief-like states should be regarded as beliefs, claiming that the distinction is 'merely terminological...nothing of metaphysical or psychological import turns on it' (Della Rocca 2003: 211). Lin and Bennett are content simply to characterize ideas as belief-like (Lin 2006: 402; Bennett 1984: 165), without sharply distinguishing beliefs from non-beliefs. By contrast, Diane Steinberg insists that even if there is no difference in kind, it is critical to Spinoza's practical epistemology that he maintains a distinction between beliefs and other ideas (D. Steinberg 2005: 156).

⁷ One could resist this conclusion by insisting that mere imaginings are beliefs, but not errors. While there is no indication that Spinoza distinguishes between false beliefs and errors, I could concede this without sacrificing much, since the puzzle that motivates this paper could be reframed in terms of dominant and non-dominant beliefs. However, I think there are good reasons to prefer a more restricted account of belief. For instance, Spinoza allows that imagining non-existent things could be a virtue if one knew [*sciret*] that such things didn't exist (E2p17s). This claim raises problems for the permissive conception of belief, since it would imply (unspinozistically) that it could be a virtue to have false beliefs, and that one could know that $\sim p$, while simultaneously *believing* that p . One avoids these infelicitous implications by ascribing to Spinoza a more restricted conception of belief.

⁸ Spinoza's own example was probably just of a boy imagining a horse, not a winged horse. The wings were added by later translators and editors, like Carl Gebhardt. For more on this, see Totaro 2015: 328.

⁹ Diane Steinberg has called this a 'default theory of belief,' according to which 'if A has an idea that p , then she will believe that p unless she has some other idea that excludes p or she perceives that her idea that p is inadequate' (D. Steinberg 2005: 148).

¹⁰ D. Steinberg 2005: 149 ff.

¹¹ As an anonymous referee pointed out, there *could* be cases of mutual exclusion which would fit with the merely logical account. I myself left open this possibility in the earlier draft by suggesting that all offsetting ideas are 'excluding' on Spinoza's account. However, when Spinoza introduces the idea of exclusion [*tollere*] (not to be confused with *secludere*, which is sometimes also translated as 'exclude') in E2p49s, he ties exclusion to *disbelief* and not to doubt; and for Spinoza disbelief in p requires belief in something, q , that offsets p . In other words, 'exclusion' might be reserved for cases of asymmetric canceling out. Moreover, even if exclusion need not be asymmetric, the fact that it *can* be is enough to show that it exclusion requires more than mere logical incompatibility.

¹² See Della Rocca 2003; Sharp 2007.

¹³ See Gendler 2003; Gendler 2008a; Gendler 2008b.

¹⁴ Gendler 2008a: 637.

¹⁵ For a sanguine interpretation of the place of fictions in Spinoza's thought, see Gatens and Lloyd 1999. While I agree with a good deal of Gatens and Lloyd's analysis, I think that they downplay the extent to which Spinoza's main concern in the *TdIE* is to expose the epistemic deficiencies of ideas of the imagination, including fictions.

¹⁶ Elsewhere in the *Ethics*, Spinoza treats fictions somewhat more loosely to include ideas of things that appear fictitious or merely imaginary to one who knows better, but which not apprehended *as* fictions—that is, as posits that are excluded by stronger ideas—by the holders of these fictions.

¹⁷ He writes: 'Although this love toward God has had no beginning (by P33), it still has all the same perfections of love, just as if it had come to be (as we feigned in P32C). There is no difference here, except that the mind has had eternally the same perfections which, *in our fiction*, now come to it' (E5p33s).

¹⁸ See Rutherford 2010.

¹⁹ As I read him, Spinoza is committed to the view that when one represents something as possible, one represents it as *not necessary*. TTP 4, 58 invites us to contrast perceiving something as necessary and perceiving it as (merely) possible. And while his definition of possibility in *Ethics* 4D4 appears compatible with metaphysical necessitarianism, Spinoza proceeds to claim that to represent something as possible is to represent it as not necessary. We see evidence for this for instance in E4p11, where he claims not merely that when one represents something as possible one does not represent it as necessary, but further that when

one represents something as possible (or contingent, for that matter), one represents it as 'not necessary' (E4p11). In virtue of our finite, limited intellects, we are bound to regard things as merely possible, even while we struggle to understand them as necessary. What TTP 4, 58 adds is that it is 'better' in some circumstances that we persist in imagining that which we know do be false, namely that things are merely possible.

²⁰ One might fruitfully compare Spinoza's views of salutary fictions with those of orthodox Calvinists, who believed that, although our fate is sealed with respect to our election (our eternal blessedness or damnation), we ought to act as if we were free and as if we could earn our own blessedness.

²¹ See Gatens 2012. On Gatens' interpretation, the exemplar of human nature (E4 Preface) is a fiction or construct of the imagination that nevertheless functions to guide to moral deliberation. See also Rosenthal 2001.

²² Just as a child might imagine that a marshmallow is a cloud so that she can restrain the desires that arise from her belief (that it is in fact a marshmallow) (see Mischel, Ebbesen and Zeiss, 1972), or as javelin thrower might pretend that she could pierce the sky so that she can launch the javelin even further, perhaps imagining that one can alter one's share of eternity could lead one to pursue knowledge with greater vigor.

²³ In contrast to fictions (as myths or parables) that are useful (e.g. pedagogically) only when positively believed by the consumers of these fictions.

²⁴ I refer to this view as *Humean* in order to avoid interpretative disputes about Hume's own account of the relationship between fiction and belief. I am simply advancing one way of reading Hume that yields an account that contrasts with Spinoza's.

²⁵ Hume 1975: 48.

²⁶ Hume 2000: 68.

²⁷ One could also cite Hume's example of the man suspended from a high tower in an iron cage. Such a person 'cannot forbear trembling...[even while he] knows himself to be perfectly secure from falling' (Hume 2000: 100–101). The 'circumstances of depth and descent strike so strongly' that they cannot be restrained by other counterposing ideas of security (Hume 2000: 101). Hume regards this as an example of unphilosophical probable reasoning, indicating that one's *belief* here is that one is *unsafe*, even if one, in some sense, knows better.

²⁸ See Lin 2006: 402. Here, 'action' is being used in the non-technical sense. Spinoza uses the term 'action' [*actio*] and the verb 'to act' [*agere*] in this non-technical sense of voluntarily producing effects, with or without the assistance of external causes. See, for instance, E2p13s; E2p35s; E3p2s; E4p59. Curley notes this in the glossary of his translation (Spinoza 1985: 624).

²⁹ In fairness to Lin, he admits that the account provided in the example is 'far from complete' (Lin 2006: 403).

³⁰ Under the attribute of thought, an affect is an idea that tracks changes in one's power of acting. The ideas of increases in one's power of action are forms of joy; the ideas of reductions of power are forms of sadness; and the conative aspect of these increases or decreases are forms of desire. Affects are doubly representational: they directly represent a change in one's power of acting, while indirectly representing some *object* associated with the production of this change. For more on Spinoza's general account of the affects, see Steinberg 2016 and LeBuffe 2009.

³¹ For an alternative account, according to which all ideas are affective, see Shapiro 2012.

³² Thanks to an anonymous referee for this suggestion.

³³ Representation is complicated matter for Spinoza. Elsewhere I have argued that he allows for at least two different kinds of representation: representation as a success concept

(*de re* representation), where one represents *x* just in case *x* obtains; and representation in a more conventional sense that admits of misrepresentation (J. Steinberg 2013). See Garrett forthcoming.

³⁴ Lest one worry that Spinoza unwarrantedly moves from the claim that we are necessarily motivated by the greater good to a conclusion about being motivated by utility, we should note that (perceived) good and (perceived) utility amount to the same thing for Spinoza. One's actual essence (E3p7) is a striving for one's own advantage [*utilitas*] (e.g., E4p20, E4p37s1 and s2.), and things are good to the extent and only to the extent that they are advantageous [*utile*] to our striving (E4d1; E4p8d).

³⁵ Compare Hobbes 1998: 78.

³⁶ Again, compare Hobbes 1996: 39

³⁷ E3p51s supports this claim about comparative evaluative judgments, since it establishes the covariation of judgments and affects. We will consider this point in more depth below.

³⁸ In fact, on the interpretation that I propose, the relationship is biconditional: one will judge that *x* is better than *y* iff one desires *x* more than *y*. However, there is an asymmetry in that desire is the ground of evaluative judgment, rather than the other way around. See J. Steinberg 2016.

³⁹ It is worth noting that the claims here resemble the principles that generate the puzzle that Donald Davidson presents concerning weakness of will, with the two premises flipped around: P1: 'If an agent wants to do *x* more than he wants to do *y* and he believes himself free to do either *x* or *y*, then he will intentionally do *x* if he does either *x* or *y* intentionally' P2: 'If an agent judges that it would be better to do *x* than to do *y*, then he wants to do *x* more than he wants to do *y*' P3: 'There are incontinent actions' (Davidson 1980: 23).

⁴⁰ Hume considers the objection that because judgments of goodness *do not* covary with sympathy and sentiments these judgments must not proceed from the sentiments: 'as this sympathy is very variable, it may be thought, that our sentiments of morals must admit of all the same variations. We sympathize more with persons contiguous to us, than with strangers....But notwithstanding this variation of our sympathy, we give the same approbation to the same moral qualities in *China* as in *England*. They appear equally virtuous, and recommend themselves equally to the esteem of a judicious spectator' (Hume 2000: 371). His reply is to acknowledge that sentiments and moral judgments do not co-vary, but to insist that this is not in any way undermine the claim of that moral judgments are derived from the sentiments. Rather, in order to 'arrive at a more *stable* judgment of things, we fix on some *steady* and *general* points of view; and always, in our thought, place ourselves in them, whatever may be our present situation' (Hume 2000: 372). He describes this as a method of 'correcting our sentiments, or at least, of correcting our language' (Hume 2000: 372).

⁴¹ This designation is owed in part to Andrew Youpa, who considers the possibility that E3p9s should be read as claiming that 'desiring *x* constitutes judging *x* good' (Youpa 2007: 384). Ultimately, Youpa rejects this reading on the grounds that E4p59 alt. dem. and E4p19 taken together show that we desire things *because* we judge them to be good, in which case the judgment is not constituted by the desire. However, if we understand the Constitution Thesis as claiming that evaluative judgments are constituted by affects in general, Youpa's concern may be countered. On this reading, evaluative judgments give rise to desires because evaluative judgments are nothing but affects, and affects give rise to desires. Moreover, as I have argued at length elsewhere (J. Steinberg 2016), the sense in

which affects—understood here as restricted to forms of joy and sadness—give rise to desires does not imply that affects precede desires. Rather desires are just the conative side of joy or sadness. See note 42 below.

⁴² The relationship between desire and affect is rather complicated. I have argued elsewhere that affects are just the ways in which one's essential striving gets constituted and particularized, such that, for instance, an instance of joy is rooted in desire (striving to increase one's power of acting) though it also orients this primordial desire to affirm what one imagines to be the cause of the joy (J. Steinberg 2016). While one can conceptually separate the joy from the desire (see, once again, Definition of Affects II, 192), instances of joy and particular desires are formed together. On my reading, then, relative judgments of goodness will track relative intensity of desire because judgments of goodness reduce to affects, and affects co-vary with desires (see E3p37).

⁴³ Indeed, it might seem to support the distinction between judgments from affects, if we understand the 'from' in the expression that 'each one, from his own affect judges...' as indicating some temporal priority of affects. However, my proposal is that we understand 'from' here as flagging explanatory priority: from the very fact that one has the affects one does, it follows that one judges such-and-such.

⁴⁴ To have a *cognitio* of *x* is not to satisfy certain epistemic conditions in relation to *x*; rather, it is simply to *cognize* or represent *x* in some way. The demonstration of E4p8 makes this clear, since the cognition in question is that of (mere) perception. And further support for this can be found in the demonstration of E4p64, which explicitly draws on E4p8 in order to show that 'cognitio of evil is an inadequate *cognitio*.'

⁴⁵ The demonstration itself is muddy, at best. Leaving aside the last clause ('insofar as we are conscious of it'), which is treated separately (for more on this, see J. Steinberg 2016), the main argument runs something like this: good and evil are nothing but increases or decreases one's power of acting, and increases and decreases in one's power of acting are forms of joy and sadness, respectively. Consequently, good and evil are nothing but forms joy or sadness. So, to represent something as good or evil is just to represent it as the source of joy or sadness. This last step is highly questionable, since it assumes that we can substitute coextensive terms in an intensional context, *salva veritate*. Without trying to vindicate what looks like unsuccessful demonstration, I will simply note that Spinoza might not actually regard this as an intensional context. To see why, one must look at his multi-layered theory of representation (see J. Steinberg 2016; Garrett).

⁴⁶ The Constitution Thesis also helps to explain his claim that 'decisions of the mind [*mentis decreta*] are nothing but the appetites themselves' (E3p2s). In this passage he attempts to dispel the notion that some of our actions can be explained only by reference to the power of the mind, showing instead that 'the decision of the mind' and the 'determination of the body' are 'one and the same thing' (E3p2s, g ii.144). The point is that decisions of the mind and determinations of the body are both fixed entirely by the dynamics of affects or appetites. There is no independent causal role for evaluative judgment. This passage can be reconciled with his later claims that we act from judgments of good or evil, if we accept the Constitution Thesis: evaluative judgments are constituted by affects; they are not acts of mind distinct from affects.

⁴⁷ See Della Rocca 2008a; Della Rocca 2008b.

⁴⁸ See H.H. Price 1964: 162.

⁴⁹ See, again, Davidson 1980: 27.

⁵⁰ Admittedly, there is something quite odd about the Constitution Thesis. It would seem that even if evaluative judgments are rooted in desires, one could recognize that in a

particular instance one's occurrent affects-desires reflect cognitive distortions (e.g. the tendency to hyperbolically discount the future), in which case one might sincerely judge that one ought to do something other than what one most desires now. For instance, I might recognize that my aversion to going to the gym is based on the saliency of the sound of the whipping wind, which leads me to exaggerate the *badness* of the walk to the gym and impairs my ability to represent the distal goods that follow physical exercise. And so I correct for this in my judgment, even if my affects don't follow suit. Spinoza's response would be to admit that the more time-neutral an evaluation is, the more rational it is (E4p62). However it is one thing to *have* a rational idea, and it is quite another thing to have one's judgment strongly informed by, or dominated, by this rational idea. If I am not very moved by rational considerations, it must be that my rational ideas only weakly inform my judgment, and that my irrational ideas, in this case the appraisal 'outside, BAD!' exert a far greater influence on my judgment, even if I do not identify with these ideas, and even if they harmonize less well with other ideas and appetites that I might have.

⁵¹ For two of the most in-depth analyses, see Lin 2006 and Marshall 2008.

⁵² Echoed in Paul's epistle: 'For though the will to do good is there, the deed is not. The good which I want to do, I fail to do; but what I do is the wrong which is against my will' (Romans 7:18–19).

⁵³ In two other instances when he is presenting *akrasia* without directly quoting Ovid, he drops '*proboque*' from the formulation, stating instead: 'we see the better and follow the worse' (E3p2s; Ep. 58 to Schuller).

⁵⁴ Lin seems to take 'acting against our better judgment' to be acting against our more rational judgment ('better judgment' in objective sense) (Lin 2006: 395, 405); but this is not *akrasia* in the strict sense. Marshall claims that Spinoza accounts for strict *akrasia*, which he understands as 'synchronic' *akrasia* (Marshall 2008: 43), or acting contrary to one's judgment of the best *at that same time*: 'the Spinozist theory explains strict akratic action, while the Davidsonian one cannot' (Marshall 2008: 54). He defends an interpretation of Spinoza according to which one can act contrary to one's rational *judgment*; but he does so only by assuming that two opposing ideas can both count as judgments: 'our partial or short-term judgement may be more effective in motivating us to act than our rational judgement. As a result, we act on a judgement about what we perceive to be good in some sense, against our judgement about what would be best' (Marshall 2008: 52). There are a couple of problems with this. First, this seems to run contrary to Spinoza's conception of belief (and judgment), as exemplified by (1). Second, even we allow for competing *judgments*, it permits us to reformulate strict *akrasia* as action contrary to one's *dominant* judgment, which still appears to be incompatible with (3). Marshall's solution would apparently be to reject (3), since he denies that judgments of the best must track desires: 'And Spinozist akratic action has a similar explanation—it is done with the belief that it is not in our best interest, yet is in fact what we most want to do' (Marshall 2008: 53).

⁵⁵ Offsetting may take the form either of opposition by ideas with countervailing content [*tollere*] or of separation in virtue the piling up of intervening ideas [*secludere*]. The former would lead to a greater erosion of confidence. See E2p17; E2p49s; E4p10d; TdIE §78.

⁵⁶ Spinoza acknowledges that constancy of judgment comes in degrees (see E3p18s1).

⁵⁷ Hobbes 1996: 44.

⁵⁸ For evidence that Spinoza allowed that the intensity of affects (i.e. evaluative judgments) could be measured diachronically, see E5p7 and E5p7d.

⁵⁹ An anonymous referee has reasonably suggested that this putative attempt to reconcile (3) and (4) looks ultimately like a rejection of (4), since (4) requires the possibility of acting contrary to one's judgment at the very time of one's decision. However, I would

maintain that this account is consistent with (4). To see this, consider the case of non-evaluative judgments. Let's say I steadily affirm that Henry Fielding is the author of *The History of Tom Jones*. However, for a very short interval (say, when my glucose is peaking after a big meal), I cannot recall the author's name, only to affirm minutes or even seconds later that it was indeed Fielding. If we ask whether I believed that Fielding is the author *during the time of lapse*, my intuition is to say 'yes.' (If the lapse were for merely a few seconds, did I really cease to believe that Fielding was the author and then come to believe this again? What about if I'm half-asleep? Or intoxicated? Do my beliefs change just because my disposition to respond is temporally altered?). A parallel case could be made vis-à-vis evaluative judgments: even at the very moment when Ulysses most wants to give himself over to the Sirens, we might say that this is not what he really judges to be best, since we can see that he has a steady preference to resist their call. If this were not the case, we would have to say that when Ulysses is bound to the mast, present-Ulysses is being held hostage by the judgment of past-Ulysses. This strikes me as a perverse description of the situation. My speculative proposal then is that Spinoza could say that Ulysses' judgment *over the entirety of some (indeterminate) stretch of time* is revealed in his steady preference, irrespective of the state of his affects at the very time-slice in question.

⁶⁰ See also E5p8 and E5p11.

⁶¹ See for instance Gordon 1995.

⁶² Compare with Pascalian feigning, where further evidential props are adopted to encourage the flourishing of one's idea of God's existence.

⁶³ Audiences at Queens University and the New England Colloquium in Early Modern Philosophy (Yale University) pushed me to clarify several key claims. Eugene Marshall provided an excellent commentary on the second part of this paper at the 2014 Pacific APA meeting. And an exacting and thoughtful anonymous referee for this journal helped me to refine the paper in numerous ways. I am grateful for all of the assistance.

REFERENCE

- Bennett, J. (1984), *A Study of Spinoza's Ethics*. Indianapolis: Hackett.
- Curley, E. (1975), 'Descartes, Spinoza and the Ethics of Belief', in M. Mandelbaum and E. Freeman (eds) *Spinoza, Essays in Interpretation*, eds. edn. Open Court: LaSalle, IL.
- Davidson, D. (1980), 'How is Weakness of Will Possible', in *Essays on Actions and Events*. Oxford: Oxford University Press.
- Della Rocca, M. (2003), 'The Power of an Idea: Spinoza's Critique of Pure Will', *Noûs*, 37(2): 200–231.
- (2008a), *Spinoza*. London: Routledge.
- (2008b), 'Rationalism Run Amok', in C. Huenemann (ed.) *Interpreting Spinoza: Critical Essays*, ed. edn. Cambridge: Cambridge University Press.
- Descartes, R. (1984), in J. Cottingham, R. Stoothoff and D. Murdoch (eds) and trans. *The Philosophical Writings of Descartes*, Vol. 2. Cambridge: Cambridge University Press.
- Garrett, D. (forthcoming), 'Representation and Misrepresentation in Spinoza's Philosophy of Mind', in M. D. Rocca (ed.) *The Oxford Handbook of Spinoza*. New York: Oxford University Press.
- Gatens, M. (2012), 'Compelling Fictions: Spinoza and George Elliot on Imagination and Belief', *European Journal of Philosophy*, 20(1): 74–90.
- Gatens, M. and Lloyd, G. (1999), *Collective Imaginings: Spinoza, Past and Present*. London: Routledge.

- Gendler, T. (2003), 'On the Relation Between Pretense and Belief', in D. Mclver Lopes and M. Kieran (eds) *Imagination Philosophy and the Arts*. London: Routledge: 125–141.
- Gendler, T. (2008a), 'Alief and Belief', *Journal of Philosophy*, 105(10): 634–663.
- (2008b), 'Alief in Action (and Reaction)', *Mind and Language*, 23(5): 552–585.
- Gordon, R. (1995), 'Sympathy, Simulation, and the Impartial Spectator', *Ethics*, 105(4): 727–742.
- Hobbes, T. (1996), in R. Tuck (ed.) *Leviathan*, ed. edn. Cambridge: Cambridge University Press.
- (1998), in R. Tuck and M. Silverthorne (eds) and trans. *On the Citizen*. Cambridge: Cambridge University Press.
- Hume, D. (1975), in L. A. Selby-Bigge and P. H. Nidditch (eds) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, eds. edn. Oxford: Clarendon Press.
- (2000), in D. F. Norton and M. J. Norton (eds) *A Treatise of Human Nature*. Oxford: Oxford University Press.
- LeBuffe, M. (2009), 'The Anatomy of the Passions', in O. Koistinen (ed.) *The Cambridge Companion to Spinoza's Ethics*. Cambridge: Cambridge University Press.
- Lin, M. (2006), 'Spinoza's Account of Akrasia', *Journal of the History of Philosophy*, 44(3): 395–414.
- Marshall, E. (2008), 'Spinoza on the Problem of Akrasia', *European Journal of Philosophy*, 18(1): 41–59.
- (2013), *The Spiritual Automaton*. Oxford: Oxford University Press.
- Mischel, W., Ebbesen, E. B. and Zeiss, A. R. (1972), 'Cognitive and Attentional Mechanisms in Delay of Gratification', *Journal of Personality and Social Psychology*, 21(2): 204–218.
- Price, H. H. (1964), 'Half-Belief', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 38: 149–162.
- Rosenthal, M. (2001), 'Tolerance as a Virtue in Spinoza's Ethics', *Journal of the History of Philosophy*, 39(4): 535–557.
- Rutherford, D. (2010), 'Spinoza's Conception of Law: Metaphysics and Ethics', in Y. Melamed and M. Rosenthal (eds) *Spinoza's Theological-Political Treatise: A Critical Guide*, eds. edn. Cambridge: Cambridge University Press.
- Shapiro, L. (2012), 'How We Experience the World: Passionate Perception in Descartes and Spinoza', in M. Pickavé and L. Shapiro (eds) *Emotion and Reason in Early Modern Philosophy*, eds. edn. Oxford: Oxford University Press.
- Sharp, H. (2007), 'The Force of Ideas in Spinoza', *Political Theory*, 35(6): 732–755.
- Spinoza, B. (1985), in E. Curley (ed.) and trans. *The Collected Works of Spinoza*, Vol. 1. Princeton: Princeton University Press.
- (2007), in M. Silverthorne and J. Israel (eds) and trans. *Theological-Political Treatise*. Cambridge: Cambridge University Press.
- Steinberg, D. (2005), 'Belief, Affirmation, and the Doctrine of *Conatus* in Spinoza', *The Southern Journal of Philosophy*, 43: 147–158.
- Steinberg, J. (2013), 'Imitation, Representation, and Humanity in Spinoza's Ethics', *Journal of the History of Philosophy*, 51(3): 383–407.
- (2016), 'Affect, Desire, and Judgement in Spinoza's Account of Motivation', *British Journal for the History of Philosophy*, 24(1): 67–87.
- Totaro, P. (2015), 'The Young Spinoza and the Vatican Manuscript', in Y. Melamed (ed.) *The Young Spinoza*. Oxford: Oxford University Press.
- Youpa, A. (2007), 'Spinoza's Theory of Motivation', *Pacific Philosophical Quarterly*, 88: 375–390.