

Reconciling Practical Knowledge with Self-Deception

(forthcoming in *Mind*)

ERIC MARCUS
Auburn University
marcuea@auburn.edu

Is it *impossible* for a person to do something intentionally without knowing that she is doing it? It might seem that the Knowledge Thesis (as I will call it) is false, or at least controversial enough to be a poor starting point for philosophical inquiry into the nature of intentional action. This is, however, where Anscombe begins *Intention*. Is this a mistake? Obviously, if the Thesis is false, then no one should accept any conclusion that she or anyone else draws from it. But if it's true, it is a remarkable fact, one that should be at the center of our thought about intentional action. This suggests that it should be among the first items to consider.

My aim in this essay is to defend the Knowledge Thesis from what I find to be among the most compelling putative counterexamples, those involving self-deception: cases in which someone is not (at least in a straightforward sense) *lying*, but yet disavows a correct description of her intentional action. Take, for a simple example, a man who deceives himself about why he finishes his younger brother's sentences. He insists that he is just being helpful, but really (we can suppose), he is doing it in order to undermine his brother. I will argue that the older brother *does*—indeed *must*—know what he's doing, his disavowals notwithstanding.

Although my primary aim is to mount an Anscombean defense of the Knowledge Thesis against a potential counterexample, this defense will also bring into view a phenomenon that has gone largely unexplored in the literature: what I will call *self-*

deceptive action. Philosophers have focused on subjects who are deceiving themselves about facts that bear on their significant concerns, e.g., whether a spouse is faithful, a child is intelligent, a course of illness will prove fatal, etc. But in cases of self-deceptive action, an agent is deceiving herself about what she is intentionally doing. This is not of merely taxonomic interest. Approaches that might be plausible in considering self-deception about facts external to the exercise of one's agency are not plausible when they are extended to facts about what one is doing. I will argue that the *deflationist* approach, for example, according to which (to put it crudely) the self-deceived subject believes what she wants to but not what she doesn't, cannot be maintained in cases in which what she doesn't want to believe is that she is intentionally pursuing an aim *that she is in fact intentionally pursuing*. Thus the reconciliation of practical knowledge with self-deception promises to shed new light on both.

My discussion is divided into three parts. I begin, in section I, by illustrating the challenge that self-deceptive action poses for an account that takes the Knowledge Thesis as fundamental. For this purpose, I focus on *Intention*. Self-deception is not, as one might initially suspect, a phenomenon that can be shrugged off as simply outside the scope of intentional action. In section II I argue that self-deception is compatible with practical knowledge. A premise of the argument is that we should understand such knowledge in the first instance as Anscombe does, in terms of the ability to *just say* what one is doing, i.e., to say so authoritatively but not on the basis of observation or evidence. I contend that the factors that explain self-deceived disavowals of action are better understood as *masking* that ability than as demonstrating its absence. And in section III, I show that the view of self-deception required by the reconciliation strategy outlined in II can be elaborated in a way that is both intuitively plausible and congenial

to recent theoretical work on self-deception. Roughly: when we are deceiving ourselves about what we are intentionally doing, we don't think about our action because it's painful to do so.

I.

An intentional action performed by an agent yet non-lyingly disavowed is a *self-deceptive action*. My question is whether an agent must know that she is performing even self-deceptive actions. This is distinct, I should emphasize, from the question of whether the process of self-deception is *itself* an intentional action.¹ These two questions are, however, linked in the following way. Part of the reason why philosophers resist conceiving of the process of self-deception itself as intentional is because it would suggest that the self-deceiver *knowingly* deceives herself.² This resistance thus has its source in part in the thought that there is a tight connection

¹ It is also distinct from the question of what Eric Funkhouser has recently called Practical Self-Deception: the question of whether values, lifestyles and other broadly practical phenomena can *embody* self-deception, in the sense that they exist as the result of deceiving oneself about what is worth desiring, valuing, or living by. See Funkhouser 2012.

² It is on these grounds, e.g., that Lynch argues that self-deception itself cannot be intentional. See Lynch 2011, ch. 5. I will argue in discussion of objection (ii) below that, in fact, KT is consistent with intentionalism about self-deception.

between intentional action and practical knowledge. And this is precisely what leads to the puzzle of self-deceptive action. We can frame this puzzle as an inconsistent triad:

The Knowledge Thesis (KT): It is impossible for a person to do something intentionally without knowing that she is doing it.

Intentionality (IN): Self-deceptive action is intentional.

Ignorance (IG): A person who is self-deceived about doing something does not know that she is doing it.

The most important contemporary discussion of KT occurs in *Intention*. It will be useful to see how this puzzle arises in the framework developed there. The difficulty, at bottom, is this: Given how Anscombe conceives the Knowledge Thesis, what might seem to be the least-painful solution to the puzzle—the rejection of IN—is not plausible. Indeed, it is implausible quite independently of her philosophical commitments. To maintain KT, IG must be rejected. But it's hard to see how to do that credibly, as IG will sound to many ears like little more than a restatement of the very idea of self-deceptive action. But I will ultimately show, on broadly Anscombean grounds, that and why we should reject IG.

Intentional actions, Anscombe famously proposes, are those to which a certain sense of the question 'why?' has application—the question that asks for the agent's

reasons for acting.³ Whether it has application is shown by whether the agent gives it application. One way of refusing to give it application is by saying “I was not aware I was doing that.” The agent might be lying, but if not then it lacks application.⁴ The account thereby entails that to be x-ing intentionally is (among other things) to be aware that one is x-ing—which for Anscombe is *knowing* that one is x-ing. Anscombe’s method—and the account she arrives at using this method—would seem to depend on there not being any cases of intentional action in which the agent non-lyingly refuses the ‘why?’ question. And yet this is exactly what happens in, for example, the case of the sentence-finisher. If asked why he is undermining his brother, he would refuse the question. Since he is self-deceived, he is not *lying* when he does so. Yet he is undermining his brother. His doing so is not an accidental byproduct of his sentence-finishing but his true aim. Nor are we simply speaking of a motive (say, cruelty) for performing the action. Rather we find in this case “the order which is there whenever actions are done with intentions” (Anscombe 2000, p. 80). He is finishing his brother’s sentences *in order to undermine him*. It would seem, then, that he is doing the latter intentionally and, if Anscombe is correct, with knowledge that he is.

There are two ultimately unpromising responses that an Anscombean might make here: one is a non-starter, the other is tricky. The non-starter is that KT is only meant to apply to *normal* cases of intentional action. According to this response, it is no threat to her view if agents are occasionally unaware of what they are doing intentionally, so long as these are non-paradigmatic instances of intentional action. But

³ Anscombe 2000, p. 11.

⁴ Anscombe 2000, p. 11.

this can't be right: her test flatly excludes the idea that there could be *any* cases in which a subject non-lyingly rejects the question but is nonetheless intentionally x-ing. For those are precisely cases in which the question that she puts forward as delimiting the scope of intentional action has been shown not to apply. And, even more importantly, the argument for her account of intentional action is predicated on the account's explaining the applicability of the 'why?' question, understood in part in terms of the agent's giving or not giving it application. It is, in other words, a premise of her argument that the events at issue are *as such* known.

The second response is to reject IN. It goes like this: Although the ordinary term 'intentional' applies in cases like that of the sentence-finisher, Anscombe's topic is not simply whatever the ordinary term picks out. If self-deceptive actions are not practically known, then they are not what her theory is meant to explain. To see why this is not naked ad hocery, we need to reflect on her method. The substance of *Intention* is an account of certain events that are known in a specific way—an explanation of what intentional actions must be given that the agent knows she is performing them, but not on the basis of observation or evidence. The account, if successful, exhibits the real unity of the class demarcated by Anscombe's various tests. The role that self-knowledge is playing in the account is not simply to specify criteria by which the relevant phenomenon might be recognized, but to say what makes the phenomenon the phenomenon it is. On her view, a particular sort of self-knowledge constitutes the intentionality of the action. Thus it is reflection on what can and can't be non-perceptually and non-evidentially known that initially points towards the idea of intentional action as sharing a structure with practical reasoning. What can be so known is constrained, she observes, by what can be thought of by the agent as realizable

by available means.⁵ This leads her to the conclusion that the entire teleological structure that a single action comprises is constituted by thought: “practical knowledge is ‘the cause of what it understands’” (Anscombe 2000, p. 87).

The metaphysical unity discerned in a category of action initially demarcated in epistemological terms would give some credibility simply to shrugging off would-be counterexamples as outside the scope of her inquiry. This would not simply be a case of conceptual gerrymandering. The Anscombean can in good conscience allow that the ordinary term ‘intentional’ applies outside the class of practically-known actions, so long as the latter class is shown to be a theoretically central restriction of the wider class of human activity. An Anscombean could then use this as a justification for concluding that, so long as ‘intentional’ is interpreted in Anscombe’s quasi-technical sense, IN is false. I call this way of resolving the puzzle of self-deceptive action *the shrugging strategy*.⁶

It will be helpful in discussing this point to have on hand a more fleshed-out example:

Maya and Gil’s recently-purchased house is surrounded by a large, unruly yard, with all manner of tangled vines, trees, shrubs, etc., occupying what has the potential to become a lovely outdoor space. As neither of them have any interest in or aptitude for such matters, they

⁵ Anscombe 2000, pp. 34-36.

⁶ Anscombe herself might have opted for the shrugging strategy: see Anscombe 2000, pp. 25-26.

agree to hire a landscaper to clear and replant it. Johnny comes highly recommended and soon proves himself to be adept at both the slow, demanding work of clearing, and the creative work of designing the look of the garden. However, soon after Gil notices that Maya and Johnny have struck up a friendship, he declares that it's foolish to pay someone else so much to do what he could do himself. Johnny is let go.

Notwithstanding the exhaustion, the injuries, the very slow pace and the homely results, Gil persists in doing the job himself—until, that is, he hears through the grapevine that Johnny has given up his landscaping business to become the manager of the garden center at a local Home Depot. Gil admits defeat and they hire a new landscaper. Thereafter, Gil—never previously known to volunteer to run errands—frequently asks Maya if they need anything from Home Depot, as he is “going to be in the area and could save her the trouble”. This bout of errand-running lasts until some months later, when Johnny moves to Miami. When Maya confronts him with the charge that he has been intentionally stopping her from seeing Johnny—by misguidedly depriving her of any occasion for doing so—he vigorously denies it. “You know that I’ve been looking for a hobby for a long time—landscaping was a great idea. It just didn’t pan out.” And: “You’re the one who always says that I never help out with the shopping!”

Gil, we will suppose, landscapes and runs errands to Home Depot in order to stop Maya and Johnny from seeing one another. We will suppose further that Gil is not *lying* when he denies doing these things for that reason.

According to the shrugging strategy, although Gil is intentionally landscaping and running errands to Home Depot and doing so in order to stop Maya from seeing Johnny, Gil is not intentionally stopping Maya from seeing Johnny. This is puzzling on its face and should be especially so for an Anscombean. For Anscombe holds that a means-action is the conclusion of a practical syllogism that has the end as a premise.⁷ The end's serving as a premise is its being intentional.⁸ So for the Anscombean to say that an end-action (stopping them) is not intentional would be to say that although the means (landscaping or errand-running) might—perhaps even must—be the conclusion of a practical syllogism, it is the conclusion of a syllogism with an entirely different premise (one that has nothing to do with stopping them). And this premise corresponds to an intentionally performed end-action for the sake of which she is performing the means. At the same time, an Anscombean would have to accept—in order to avoid simply denying the phenomenon—that the means-action is *also* performed for the sake

⁷ See Anscombe 2000, pp. 60-61, 65.

⁸ When I speak of 'the means action' and 'the end action', I don't mean to suggest that for Anscombe these must be distinct events. If I turn on the light by flipping the switch, 'flipping the switch' and 'turning on the light' are two descriptions of the same event. But for the event to be intentional under the description 'turning on the light' is for that end to be the premise in the practical syllogism that has 'so I'll flip the switch' as the conclusion.

of the suppressed end, which does *not* correspond to the premise in a practical syllogism, and whose teleological connection to the means-action would require an entirely different account.

Although this maneuver promises an escape from the puzzle of self-deceptive action, it comes at a high cost. For Anscombe would then have to postulate a kind of practical teleology exactly like that of intentional action save the absence of practical knowledge. “[T]he order which is there whenever actions are done with intentions” is present, but Anscombe’s ‘why’ question does not apply. How, then, could this order be constituted by the subject’s practical knowledge of what she’s doing? If there is a non-knowledge-requiring connection between the means and end of a self-deceptive action, this connection will then inevitably appear to be the essence of the teleology of human action quite generally, with practical knowledge a mere contingent add-on. The rejection of IN would thus backfire disastrously.

Even if we put aside any commitment to Anscombe’s philosophy of action, the rejection of IN is implausible. Plainly, self-deceptive actions are not *unintentional* in the ordinary sense. In that sense, roughly, what I do unintentionally is an accidental byproduct of something I’m doing intentionally. However this rough characterization is to be elucidated, it will not cover a case in which x-ing is the point of my action.

This last contention might be doubted, however. It might be thought that someone who has an implicit bias—to the effect, say, that women can’t do logic as well as men—might perform a variety of actions organized around the end of preventing a woman from being hired to fill a logic position. Nonetheless, it might be held that he is not doing this intentionally. Perhaps, then, we should treat self-deceptive action the

same way. Such behavior is not *unintentional* action as ordinarily conceived; but it is also not intentional action.

To fully reply to this objection would require more discussion of implicit bias than I can manage in the confines of this essay. But let us ask: Why does it seem plausible (if it does seem plausible) to say that the search committee member is not intentionally preventing a woman from getting the job? The answer, I would think, is this: preventing a woman from getting the job is not genuinely his aim. This is puzzling to be sure. The agent's behavior seems designed to bring about a certain end, but without this end being his aim. Unlike in the case of self-deceptive action, someone who operates under the influence of implicit bias can honestly—and not just non-lyingly—deny that he is performing the relevant action. If he is confronted with decisive evidence of his bias, this can come as a revelation. But the puzzle of self-deceptive action lies elsewhere: Someone who deceives himself about his x-ing *really does aim to x*. And this is why his disavowal of x-ing is *not* honest, even though it is also not a lie. And it is why he can come to know what he is doing in a manner that is distinctively first-personal: he can come to acknowledge not just that he is x-ing but that he knew it all along. I will return to this point below.

I take it, then, that neither Anscombe nor anyone else should attempt to resolve the paradox of self-deceptive action by rejecting IN. To hold on to KT, we must reject IG.

II.

Maya asks: “Why are you keeping me from seeing Johnny?”. Gil non-lyingly replies “I am not doing that”. Supposing that this action is intentional, many will take it to be a counterexample to KT. Why does it seem this way? Because if Gil knows that he is keeping them separated, then he *must* be lying were he to deny it. The primary burden of someone who would reconcile practical knowledge with self-deception is to refute this “must”. I will do that in this section.

Practical knowledge, as Anscombe characterizes it, is a species of “being able to say” (Anscombe 2000, p. 14), and the sort of knowledge that specifically characterizes intentional action is a being able to say not on the basis of observation or evidence, i.e., a being able to *just say* what one is doing. I take it that this much is not controversial. It is sufficient for knowing what one is doing that one can just say what one is doing in this sense. So let us ask: Is Gil able to just say that he is stopping Maya and Johnny from seeing each other?

This is certainly an odd question, one that we never have an occasion to ask in ordinary circumstances. And some might take it to be part of the description of the case that Gil is *not* able to just say so, and that to think otherwise is to render him indistinguishable from the bald-faced liar. But that’s wrong. It may be part of the description of the case that he *can’t* just say what he’s doing. But at issue is whether he has the ability, and this is a different question. For it is not obvious—and not true, I will argue—that the sense of ‘can’t’ that figures in the statement “Gil can’t just say what he is doing” is one that signifies the absence of an ability. The absence of an ability, after all, is but one explanation among others of why someone can’t do something.

To see this, note that whereas possession of the ability to x entails a non-accidental connection between an attempt to x and x-ing itself, this connection is not a necessary one. It is not, in other words, impossible that someone able to x, and given the opportunity, fails to do so when she tries. Something has interfered to stop the ability from being exercised.⁹ These interfering factors *mask* the presence of the ability. A man might have the ability to juggle six flaming torches, but it is masked by a jerk in the audience who keeps shooting rubber bands at his face, or by his drunkenness, or by a temporary drop in self-confidence due to a recent divorce. None of these conditions destroy the ability, they just make it impossible for him to exercise at the moment. In these circumstances, he cannot juggle six flaming torches. Nonetheless, he has the ability to do so. (Note that masks can be either external or internal impediments to exercising an ability.) Thus, it doesn't follow from the fact that an agent can't x that she lacks the ability to x.

Since it is possible to have an ability that one cannot currently exercise, the mere fact that a self-deceived person cannot just say what she is doing does not by itself refute KT. To do that, we would need some reason for thinking that, in cases of self-deception, the relevant 'can't' is best explained by inability.

Let's go back to Gil. Suppose it is too painful for him to acknowledge what intentionally stopping Maya and Johnny from seeing each other reveals about him: that he is insecure about Maya, that he worries about an affair, about being abandoned. He can't just say what he's doing. But is this "can't" an "unable to"? What is the correct interpretation of this complex of pain, insecurity and fear? Does it negate or merely

⁹ I discuss abilities in more detail in Marcus 2012, ch.2.

mask the ability to just say what he is doing? I will argue in the next section that the latter answer is grounded in a plausible characterization of self-deception.

But before turning to self-deception itself, it will be useful to bolster the ‘mask’ interpretation in a different way: by directly providing a rationale for thinking that even the agents of self-deceptive actions must know what they’re doing. Ultimately, the argument to follow is just a way of articulating the conception of intentional action that animates Anscombe and those drawn to her approach to understanding it. It begins with the following premise:

Means Recognition (MR): For an agent to be y-ing intentionally because she is x-ing intentionally (in the sense that implies y-ing in order to x): (i) the agent must know or believe that y-ing is a means of x-ing. (ii) the agent must be y-ing in light of this instrumental connection.¹⁰

According to this principle, to see Gil as stopping Maya and Johnny from seeing each other by taking over the landscaping is in part to see him as acting with the understanding that doing the latter is a way of doing the former. (I will stipulate that for the remainder of this discussion ‘understand’, ‘grasp’ and ‘take’ are neutral between epistemic and merely doxastic readings.) If he lacks this understanding or if it is in no way connected to why he did the landscaping, then his keeping them apart is merely an unintentional byproduct of landscaping.¹¹

¹⁰ See Anscombe 2000, p. 50.

¹¹ Cf., Anscombe 2000, pp. 79-80.

Why should a grasp of this instrumental connection be required? One might answer this question as follows: When one conceives a person as y-ing in order to x, one sees her as viewing y-ing as answering the ‘how shall I x?’ question. And part of what it is for her to see this *as* an answer is to take there to be an instrumental connection between y-ing and x-ing. Furthermore, ‘how shall I x?’ is only a question for her if she already takes herself to be x-ing. To say that S is y-ing in order to x is to portray her as already understanding herself as engaged in the project of x-ing, and adopting the means y-ing for the reason that it serves what she already understands to be her end. The fact that she understands herself as x-ing is then what explains why, if she suddenly realizes that she’s not succeeding in x-ing by y-ing—i.e., if she stops believing that her y-ing is a means of x-ing—then she will stop y-ing. The realization directly bears on what she already understands herself to be doing. She might then start z-ing in light of her instrumental belief in the connection between z-ing and x-ing. To revert to our example, Gil stops landscaping when he hears that Johnny took the job at Home Depot, and thereupon becomes very generous with his offers to run errands there. After Johnny moves to Miami, these offers peter out. Even in cases of self-deception, the agent persists in the relevant means-action only insofar as she takes the means action to be realizing the desired end, and therefore only in virtue of the fact that she understands herself as pursuing that end.¹²

¹² This rationale for KT ties it to *non-basic* action: one must know that one is x-ing insofar as to be x-ing is to be practically oriented towards x-ing as an end. This would appear to leave open whether the mask-interpretation is equally well motivated for teleologically basic actions: actions that one performs intentionally but not by

This argument is based on a widely but not universally accepted premise.¹³ It has the form of an inference to the best explanation, but it doesn't consider competing explanations. And it is neutral between requiring belief and knowledge of what one is intentionally doing. To remedy these defects would be to transform it into a full-blown defense of the Knowledge Thesis, which would require an essay of its own. Nonetheless it does, I hope, formulate in an intuitive fashion a positive conception of intentional action—including self-deceptive action—that ties it inextricably to practical knowledge, thereby motivating the 'mask' interpretation of self-deception.

This understanding of KT's rationale also helps to rule out an alternative reply to the puzzle of self-deceptive action. According to this response, Gil might exploit uncertainty about the effects of his taking over the landscaping in order to non-lyingly disavow *knowledge* of separating them. "Who's to say this will separate them?", he might rhetorically ask himself. But although it's plausible that this maneuver might play some role in self-deceptive action, it cannot account for the central element of the phenomenon: the suppression of the agent's knowledge of her end—the end conduciveness to which makes further actions worthy of choice. Gil is deceived about

performing other actions. However, I am (under the influence of Lavin 2013) doubtful that there are any such actions. And even if there were teleologically basic actions, it is not obvious that we could be self-deceived about them. I perform a self-deceptive action when I reject a *further* description of what I am doing in doing something else that I do not deny that I'm doing. I can think of no case, either discussed in the literature or of my own devising, that does not have this structure.

¹³ Clarke 2010, e.g., takes himself to have counterexamples to a principle similar to MR.

not just what he is actually accomplishing but also, and centrally, what he aims to accomplish in adopting various means to that end. In other words, insofar as he finds it painful to acknowledge that separating them is something that he's intentionally doing, it is less because it is something that he manages to accomplish than because it is something that he is taking steps in order to accomplish.¹⁴

III.

So far we have seen, in a skeletal fashion, how self-deception can fit with practical knowledge: self-deception masks it. But for this to qualify as a reconciliation, we need some understanding of the form that this masking takes. The proposal to follow is based on the observation that knowledge does not require continuous conscious reflection on the known fact. Thus insofar as keeping a thought out of consciousness is an element of self-deception, it could explain how someone who non-lyingly denies a certain fact might nonetheless know it to obtain. My goal in this section is modest: not to endorse a specific theory of self-deception, but to show how an uncontroversial feature of the phenomenon can function to mask practical knowledge.

There are, of course, limits to my ecumenism. According to many, someone who is self-deceived about $\sim p$ does not genuinely hold the belief that $\sim p$ —what I will call ‘the displeasing belief’.¹⁵ Alfred Mele, for example, argues that various difficulties that beset

¹⁴Thanks to an editor for bringing this alternative to my attention.

¹⁵ See Mele 2001, Barnes 1997, and Talbot 1995. Van Leeuwen calls these ‘one belief’ views and comes close to endorsing a version himself. See his 2007.

the very idea of self-deception result from certain unwarranted assumptions. Crucial among these is “that all self-deceivers know or believe truly that $\sim p$ while (or before) causing themselves to believe that p ”—where p is the pleasing belief.¹⁶ His project is to account for the way a desire that p can, through sub-intentional mechanisms, cause someone to believe that p *without the subject’s ever believing, let alone knowing, that $\sim p$* . This combination of displeasing-belief-rejection and anti-intentionalism marks the deflationist approach to self-deception.¹⁷

Deflationism has been aided by the failure to consider self-deceptive action. If the issue is whether a spouse is unfaithful, it is perhaps not hard to imagine that the subject is simply swept towards the belief he prefers—she’s faithful—by various mechanisms that operate outside the purview of his agency. If the issue instead is whether he is persuading his wife that she is insane, he would be deceiving himself about what he is doing *while he is intentionally doing it*. But he cannot be swept so far away from understanding himself as performing the action that it is no longer in any meaningful sense an act of his own agency. Our ongoing intentional activity functions as an anchor in the perspective from which the unpleasant truth is impossible fully to obscure. Deflationism about self-deceptive action would require that even as one intentionally took measures to realize an end, altering one’s course to suit that end’s evolving requirements, one had no more idea that one was pursuing it than if it were not one’s end at all, but only an accidental by-product of one’s action. That is untenable. Deflationism is wrong about self-deceptive action, and this lack of generality provides

¹⁶ Mele 2001, p. 60.

¹⁷ See also Barnes 1997.

grounds for a more wide-ranging suspicion about the approach. I return to this issue in the conclusion.

Here is my modest elaboration on the idea that self-deception masks practical knowledge: In exercising the ability, partly constitutive of practical knowledge, to just say what one is doing, one is bringing what one is doing to consciousness—thinking about it, in the colloquial sense. In some cases—e.g., where the thought of one’s doing it is shameful to one, offends one’s pride, undermines a cherished self-conception—this bringing-into-consciousness will be painful. The pain might lead one simply to stop performing the action. But it might also cause one to stop thinking the painful thought, thereby banishing one’s knowledge from consciousness. This is *not* the same as destroying one’s knowledge. For knowing that *p* does not require continuous conscious reflection on the fact that *p*. Knowledge—whether construed in terms of safety, epistemic virtue, reliability, rational abilities or any other of the approaches on offer—is not, nor does it depend upon, some specific continuous conscious state or activity. My proposal is that the relevant pain inhibits the ability to just say what one is doing insofar as the exercise of that ability *would* bring this thought to consciousness. Thus, the painfulness of acknowledging something (or what makes this acknowledgment painful) functions as a mask of one’s practical knowledge.

The idea that the unpleasantness associated with certain states of affairs plays a role in self-deception is perfectly familiar. We could not understand someone who was deceiving himself about facts that are emotionally neutral—say, whether the number of

hairs on his head was divisible by 47. And, aside from ‘twisted’ cases,¹⁸ we see self-deception as a response to the prospect of something deeply disturbing. This fact is also widely recognized in theorizing on the topic by philosophers across the spectrum of views of the phenomenon. Audi and Bach both describe the self-deceiver as someone who “cannot bear” the thought about which he is deceived (Bach 1981, p. 366; Audi 1988, p. 97). Johnston, Barnes and Archer each place the experience of anxiety at the center of what moves a subject to deceive themselves.¹⁹ Davidson speaks of the pain of believing $\sim p$ as motivating the self-deceptive sub-agent to promote the belief that p .²⁰ And Finagrette describes the relevant facts as ones with the potential to “gravely disrupt” the subject’s “mental equilibrium” and as possessing “traumatic significance” (Fingarette 1998, p. 295).

The datum my proposal recruits is simply that the self-deceiver finds the thought of a certain proposition’s being true to be painful. I add to this uncontroversial observation that this pain leads the subject to avoid the painful thought, and that in certain circumstances this thought-avoidance masks, without destroying, the practical knowledge that an agent who performs an action intentionally must possess. It masks this knowledge insofar as it stops the agent from being able to ‘just say’ what she’s doing by keeping her from thinking about the fact that she’s doing it.

¹⁸ In cases of twisted self-deception the subject is deceived about a fact that would, on the surface, assuage a painful thought. For an approach to twisted self-deception that could be exploited to assimilate it to the rubric I develop here, see Barnes 1997.

¹⁹ Archer 2013, Barnes 1997, Johnston 1995.

²⁰ Davidson 1985, pp. 145-146.

To go beyond this and say *how* it keeps her from doing so would require wading into the thicket of issues that has occupied the literature. I take it to be a virtue of my proposal that it sheds light on our topic without requiring us to settle the central, pre-existing questions about how self-deception works. Some view the unpleasantness associated with $\sim p$ as causing the subject to avoid bringing the belief that $\sim p$ to mind,²¹ or as helping to rationally explain the subject's directing his attention away from the displeasing fact.²² Some hold that it helps to explain both why the belief that $\sim p$ is unconscious and why the subject avows the belief that p .²³ Each of these views in effect incorporate some version of my 'uncontroversial observation' into their theory of what self-deception as such accomplishes. Other theories of self-deception, while appealing in some form to the idea that a painful thought triggers it, do not invoke consciousness or a like notion in their characterization of what Van Leeuwen calls 'the product' of self-deception.²⁴ Some view the unpleasantness associated with $\sim p$ as motivating the person²⁵ or a sub-agent of the person²⁶ to adopt the belief that p , or as an ordinary but sub-intentional cause of believing p .²⁷ These views are nonetheless compatible with my

²¹ Bach 1981.

²² Finagrette 1998.

²³ Audi 1988, Rey 1988.

²⁴ Van Leeuwen 2017.

²⁵ Talbott 1995.

²⁶ Pears 1984.

²⁷ This is one way of interpreting Barnes' thesis that self-deception is caused by 'anxious desire'. See Barnes 1997.

strategy for reconciling self-deception with practical knowledge. The fact (if it is a fact) that self-deception does not *consist* in the avoidance of painful thoughts does not rule out that it does (along the way, as it were) sometimes cause the avoidance of such thoughts, and in doing so mask practical knowledge. My proposal requires that self-deceptive action is practically known by the agent, but it is otherwise neutral with regard to questions of the intentionality of self-deception, whether it requires dividing the self into mini-selves, and what the product of self-deception is.

It might be worried, however, that in leaving so many of the central issues unaddressed, very little has in fact been accomplished, that I have done little more than reframe the mystery of self-deception. In fact, quite a bit has been accomplished.

There is a *prima facie* tension between the Knowledge Thesis and a natural interpretation of self-deceptive action: that it is performed intentionally yet without the agent's knowledge. It has been demonstrated that there is no inconsistency. This does not render the phenomenon completely unpuzzling—nor is this obviously a desideratum of a theory of self-deception—but it does help us see how self-deceptive action is so much as possible. Furthermore, what we have discovered is not simply an unoccupied but otherwise unappealing place in logical space. My proposal is motivated by a plausible conception of intentional action and can be fleshed out in a way that is congenial both to pre- and post-theoretical conceptions of self-deception. Furthermore, it bears interestingly on the nature of self-deception itself, and not simply by limiting the purview of Deflationism. I shall offer by way of conclusion a thought on how it might shape our general conception of self-deception going forward.

The view I've proposed helps also to explain the peculiar dual sense of discovery and recognition in a subject who 'recovers' from self-deception. The bringing-to-mind

of p registers as a kind of discovery, but not so much a discovery that p is true as that one already knew that p is true. One discovers that one was not conscious of what one knew all along, that one had somehow managed to keep one's own knowledge from oneself.²⁸ Relatedly, it explains why we take the self-deceived to be in a special position to confirm whether or not the hypothesis of self-deception is correct. If we speculate that Gil is intentionally separating Maya and Johnny, we think he is able to tell us—and not just because he is privy to evidence that others are not. We are looking to him not for agreement as to what the evidence suggests, but self-recognition. This could plausibly be understood as the lifting of the mask, i.e., an exercise of the ability to just say what he is doing, which is a way of imputing the action to his own agency that is in principle impossible for anyone else.

Before concluding, I will consider four objections.

(i) Deflationists tout as a crucial benefit the dissolution of the 'static paradox' of self-deception. A view that commits us to holding that the self-deceiver believes (qua deceiver) that p is false and also believes (qua deceived) that p is true saddles us with a very puzzling, if not inconceivable picture of the phenomenon. On the deflationist view the self-deceiver does not believe the proposition that he prefers to be false, so the problem simply does not arise. But the view defended here seems to require that, to revert to our example, Gil both believe that he is and also that he is not separating Maya and Johnny.

There is, however, nothing paradoxical about this. For it is not impossible that someone holds a pair of contradictory beliefs at once, only that he holds a pair of

²⁸ Thanks to a referee for help in making this point.

contradictory beliefs *in mind* at once.²⁹ To put it slightly differently, it would be impossible to conceive that S believes p and that S believes that not-p and also that nothing explains S's failure to *put them together* and see their self-evident incompatibility. For our purposes here, we can understand the senses of 'in mind' and 'together' simply in terms of the explanations that make holding contradictory beliefs intelligible: distraction, confusion, incomprehension, and repression, among others. What is impossible is for someone to hold the beliefs p and that \sim p clearly and distinctly. And beliefs that are defectively held in one of the above ways are precisely not held clearly and distinctly. Self-deception is one such non-ideal doxastic circumstance. It is a condition in which practical knowledge is pushed out of consciousness—out of one's thoughts, in the ordinary sense—thereby making it possible to believe what contradicts it.

(ii) According to a second objection, self-deceptive action involves the *intentional* masking of one's practical knowledge. Gil, it might be argued, knows that he would have to stop separating Maya and Johnny if he were fully conscious of what he was up to, and so he pushes this knowledge from his mind. He knows, that is, that consciousness of the real point of his landscaping would bring home the utter hopelessness of attempting to

²⁹ Cf., Hamlyn 1971 and Haight 1985. It might be objected that this response to the static paradox shows that my reconciliation in the end relies on a division theory of self-deception. But it does not rely on the idea of sub-agents a la Pears. I do not need to go much farther than Davidson does when he says: "I spoke of the mind as being partitioned, meaning no more than that a metaphorical wall separated the beliefs which, allowed into consciousness together, would destroy at least one" (Davidson 2004, 220).

save a marriage by such means, and that this in turn would lead him to stop landscaping. But if he were to stop, then perhaps Maya would insist on rehiring Johnny. And this would be intolerable. Separating them, he realizes, thus requires keeping the fact that he is separating them out of his thoughts. But if the agent must intentionally push her practical knowledge out of mind in order to perform a self-deceptive action, then the self-deception is itself intentional and known to the agent.³⁰

My proposal commits me to the banishment from Gil's mind of his belief that he is separating them. Whether he intentionally banishes it or whether the banishment is a quasi-mechanical effect of something else is a question to be settled by the sort of comprehensive account of self-deception from which I have prescind. Still, my thesis is *consistent* with intentionalism about self-deception itself, and hence with the interpretation of Gil on which objection (ii) is based. It would indeed follow from what I've said that the agent would have to know that he is masking his practical knowledge. But since my thesis is precisely that one can know what one is doing without having it in mind, I can allow that he also masks his knowledge that he is masking. One can intentionally evade contemplating an unpleasant fact (and so know that one is doing it) without bringing the fact that one is evading it to consciousness.

(iii) The third objection is based on a variation of the principle, discussed in section II, that y-ing intentionally in order to x entails knowledge that y-ing is a means of x-ing. According to the objection, one must *see* (i.e., have unmasked knowledge of the fact) that x-ing would, in the here and now, be an upshot of my y-ing. But if I am

³⁰ Thanks to the editors and referees for calling objections (ii)-(iv) to my attention.

conscious of this causal fact, then my knowledge that I am intentionally x-ing must itself be conscious, and hence in mind after all.

But even if we grant (for the moment) that the agent must see that y-ing would promote x-ing, this objection fails. To see why, consider a different example. Suppose that I am playing some music loudly and I know that in doing so I am annoying my neighbors. We can nonetheless suppose further that I am not annoying my neighbors intentionally. For I'm playing the loud music despite the fact it is annoying them, and just because playing loud music is the only way to shake the bomb off the bathroom sink and into the tub full of water. In this case, I am fully conscious of the fact that annoying the neighbors is a consequence of something I'm intentionally doing, but I can honestly deny that I'm annoying them intentionally. (Anscombe calls such actions voluntary, but not intentional.³¹) Gil's self-deception, we can imagine, involves taking himself to be in a position structurally analogous to that of the bomb-douser. So even if he is conscious of the fact that he is keeping them separated, his knowledge that it is *for the sake of that end* that he performs the means-action can nonetheless be masked, and thus so can his knowledge that he is *intentionally* performing the end-action. He sees that y-ing will promote x-ing, but he does not see that he is y-ing in order to x.

(iv) This reply to the third objection gives rise to a fourth. It might be contended that there are cases in which one is deceived not merely about what one is doing *intentionally*, but about what one is doing full-stop. Recall the sentence-finisher. What he denies is not simply that he *intends* to undermine his brother, but that his brother is being undermined. He could not, after all, plausibly represent his sentence-finishing as

³¹ Anscombe 2000, pp. 89-90.

helping his brother while acknowledging that it undermined him. Yet according to my concession in reply to the second objection, since he is finishing his brother's sentences in order to undermine him, he must *see* that his brother is being undermined. But then his being deceived about it cannot be a matter of not having it in mind.

Let us reexamine what was previously conceded. Why does it seem that one must be conscious of the relevant causal facts linking one's means and end? Chiefly, I suggest, for this reason: insofar as an agent views an action she is performing as to-be-done only insofar as it is conducive to a specific end, the underlying causal connection between that action and the end will tend to remain salient for her. The end, one might say, is *actively guiding* the agent. But not all y-ing in order to x involves being actively guided by x-ing in this sense. The sentence-finisher, we can suppose, is actively guided by the requirements of finishing specific sentences and, at greater remove, by the requirements of being helpful (as he interprets them). Thus his knowledge of the damaging effect of his behavior can remain masked.

This is not to deny, of course, that the older brother must at some point have consciously made the connection between sentence-finishing and undermining. We can imagine that once, when he finished his brother's sentence genuinely in order to help him, he noticed a look of humiliation on his brother's face. The reaction pleased him, and he thereafter sought to repeat the episode, now under the cover of the formerly-true story about his own motivations. This is possible in part because the cover story provides guidance in the here and now that permits the continue suppression of his knowledge that his action is having the damaging effect and is performed with the aim of producing it. This is how self-deception can mask not simply one's knowledge that one is x-ing intentionally, but that x-ing is a consequence of what one is doing.

My goal in this section has been to flesh out the idea that self-deception masks without destroying an agent's practical knowledge. To do this, I have appealed to an element of self-deception that is widely recognized and recruited it to explain the relevant masking: in deceiving oneself about what one is intentionally doing, a painful thought is pushed out of one's mind. It would require more argument to demonstrate that we ought to think of self-deception about other matters along the same lines. But it would be surprising if one found a phenomenon of this shape only in the case of beliefs on this one topic. One might think that not only beliefs about one's intentional actions, but also beliefs about one's emotions and beliefs about one's beliefs themselves might belong to this same family: even when we are deceived about them, we still possess the relevant (higher-order) beliefs, perhaps even the corresponding self-knowledge.

Suppose—admittedly without a shred of argument here³²—that this last speculation were correct. What action, belief and feeling all (arguably) have in common is this: there is some kind of constitutive connection between acting, believing and feeling, and knowing oneself to act, believe and feel in a distinctively first-personal and so non-evidential way. One might call this knowledge *self-conscious knowledge*. We could then say that self-deception—of this family anyhow—amounts to a non-lying denial of self-conscious knowledge. If that's right, then perhaps the fundamental puzzle of self-deception is not, as it is generally held, the fact that one believes against the evidence, but that one is alienated from one's self-conscious knowledge.³³

³² In fact, I have defended such a view regarding our knowledge of our own beliefs in in Marcus 2016.

³³ This paper was presented at a meeting of the Action Network at the University of

References

Anscombe, G.E.M. 2000, *Intention* (Cambridge, MA.: Harvard University Press)

Archer, Sophie 2013, 'Nondoxasticism about Self-Deception', *Dialectica* 67

Audi, Robert 1988, 'Self-deception, rationalization, and reasons for acting' in Brian P. McLaughlin & Amelie O. Rorty (eds.), *Perspectives on Self-Deception* (University of California Press)

Bach, Kent 1981, 'An analysis of self-deception', *Philosophy and Phenomenological Research* 41

Barnes, Annette 1997, *Seeing Through Self-Deception* (Cambridge: Cambridge University Press)

Clarke, Randolph 2010, 'Skilled activity and the causal theory of action', *Philosophy and Phenomenological Research* 80

Leipzig and also at Ryerson University. Thanks to those audiences and especially to Phil Clark, Jim Conant, Carl Craver, Arata Hamawaki, Boris Hennig, David Hunter, Ram Neta, Guy Rohrbaugh, John Schwenkler, Will Small, Sergio Tenenbaum, and the referees and editors at *Mind*.

Davidson, Donald 1985, 'Deception and Division', in Ernie LePore and Brian McLaughlin (eds.), *Actions and Events* (New York: Basil Blackwell)
——— 2004, *Problems of Rationality* (Oxford: Clarendon Press)

Fingarette, Herbert 1998, 'Self-deception needs no explaining', *Philosophical Quarterly* 48

Funkouser, Eric 2012, 'Practical Self-Deception', *Humana Mente Journal* 20

Haight, M.R. 1985, 'Tales from a Black Box', in M.W. Martin (ed.), *Self-Deception and Self-Understanding* (Lawrence, Kansas: University Press of Kansas)

Hales, Steven D. 1994, 'Self-deception and belief attribution', *Synthese* 101

Hamlyn, D.W. 1971, 'Self-Deception', *Proceedings of the Aristotelian Society* (supplementary volume) 35

Johnston, Mark 1995, 'Self-deception and the nature of mind', in Cynthia Macdonald (ed.), *Philosophy of Psychology: Debates on Psychological Explanation* (Cambridge: Blackwell)

Lavin, Douglas 2013, 'Must There Be Basic Action?', *Noûs* 47

Lynch, Kevin 2011, *A Defense of a Deflationary Theory of Self-Deception*, Dissertation
(University of Warwick)

Marcus, Eric 2012, *Rational Causation* (Cambridge, MA: Harvard University Press)
———2016, 'To Believe is to Know that You Believe', *Dialectica* 70

Mele, Alfred 2001, *Self-Deception Unmasked*, (Princeton: Princeton University Press)

Pears, David 1984, *Motivated Irrationality*, (New York: Oxford University Press)

Rey, Georges 1988, 'Toward a computational account of Akrasia and self-deception', in
Brian P. McLaughlin & Amelie O. Rorty (eds.), *Perspectives on Self-Deception*.
(University of California Press)

Talbott, William J. 1995, 'Intentional self-deception in a single coherent
self', *Philosophy and Phenomenological Research* 55

Van Leeuwen, D.S.N. 2007, 'The product of self-deception', *Erkenntnis* 67