Unwitting Pretense and the Self-Deceptive Mind

Abstract:

What might it mean to pretend unwittingly? I propose an answer to this question by first offering an account of the mental structures and metacognitive capacities required for imaginative pretense. I then provide an illustration of the irrationality symptomatic of the breakdown of these structures and the diminution of these capacities. The resultant state of mind I dub "unwitting pretense". In the second half of the paper I deploy the concept of "unwitting pretense" to give an account of the incongruous behavioral tendencies of subjects are who self-deceived. I argue that characterizing self-deceivers as unwitting pretenders allows us to make sense of their tendencies to avoid evidence, to subvert their own aims, and to emerge from self-deception in response to altered incentives.

Keywords:

Pretense, imagination action explanation, self-deception, metacognition

Unwitting Pretense and the Self-Deceptive Mind

Oh-oh, yes I'm the great pretender
Adrift in a world of my own

- *The Great Pretender* (Buck Ram)

Though the very idea might seem very strange, there are times when you are pretending unwittingly. In what follows, I make the case for this claim by first presenting an account of the mental structures and metacognitive capacities necessary for imaginative pretense. I then provide an illustration of the irrationality symptomatic of the breakdown of these structures and the diminution of these capacities. I dub the resultant activity, "unwitting pretense". In the second half of the paper I deploy the concept of "unwitting pretense" to give an account of the incongruous behavioral tendencies of subjects are who are commonly described as self-deceived. I argue that characterizing self-deceivers as unwitting pretenders allows us to make sense of their tendency to avoid evidence, to subvert their own aims, and to emerge from self-deception in response to changed incentives.

§1 The Role of Imagination in the Rationalization of Action

Typical human cognizers are able to represent the way the world is or could be in a variety of ways, belief being just one of these ways. David Hume displayed an early and acute awareness of this feature of mental life, making the problem of distinguishing belief from other attitudes such as imagination particularly urgent for him. Hume labored to give an account of the distinctiveness of belief in terms of its phenomenology: "This different feeling I endeavour to explain by calling it a superior *force*, or *vivacity*, or *solidity*, or *firmness*, or *steadiness*. This variety of terms, which may seem so unphilosophical, is intended only to express an act of the mind, which renders realities more present to us than fictions, causes them to weigh more in thought, and gives them a superior influence on the passions and the imagination." Ultimately, however, Hume confesses that "tis impossible to explain perfectly this feeling or manner of conception." But Hume did not rest his account on phenomenology alone. He observes a second way to distinguish belief, namely, that beliefs and not imaginings serve as "the governing principles all of our actions." (629) Beliefs, and not imaginations, have a role to play in the motivation of action.

In the twentieth-century, "dispositional" accounts of belief rejected Hume's view that belief could be distinguished by its phenomenological character, but retained Hume's view that beliefs were unique in their motivational role. Belief was thought to be distinguishable from other cognitive attitudes on the basis of its tendency to lead, in conjunction with desire, to a disposition towards action.[1] Believing that the kitchen is on

[1] cf. Price 1960/1969, *passim*)

fire motivates me to dial 911; merely so imagining, under normal circumstances, does not.

In this century, some philosophers have expressed skepticism about even the dispositional part of Hume's story. Tamar Gendler (2007), for example, maintains that any attempt to identify belief by its characteristic "marks" is bound to neglect "the numerous ways in which belief can obtain without its normal manifestations, and the numerous conditions under which other cognitive attitudes can bear them in its stead." (236) Gendler maintains that under the right conditions, imagining that *p* can come to play what is a typically belief-like role with regard to action-guidance.

But can belief and imagination really share the same kind of motivational role? In "The Aim of Belief" (2000) David Velleman proposes that imaginings motivate action in relation to desire in much the same manner as beliefs. Velleman's examples include a child imagining herself to be an elephant, a person muttering to herself during an imagined conversation with someone else, and examples from Freud in which people behave toward symbolic objects as they wish to behave toward the things that they imaginatively stand for (269). Velleman, following Hursthouse (1991), maintains that these kinds of actions stubbornly elude standard belief-desire explanation.

Velleman argues that any adequate action explanation will make essential appeal to the motivating force of imagination. Such actions, he maintains, would be utterly inscrutable otherwise. Take Hursthouse's case of "muttering imprecations under one's breath". To explain such an action without appeal to the imagination, we would have to resort to reasons that are simply implausible:

> There is nothing that we both want to do and believe ourselves to be doing by talking to ourselves in this way. If someone stopped us on the street and asked "Why were you just muttering and shaking your head like that?" we could not offer an answer that began with the words "I wanted..." What could we have wanted? To walk along muttering and shaking our heads? Hardly. (2000, 264)

Velleman proposes that in order to explain such actions we must suppose that imagining that *p* and believing that *p* are "alike in disposing the subject to what would satisfy his conations if *p* were true, other things being equal." (271) However, as Velleman realizes, things are rarely equal, and imagining that *p* and believing that *p* typically dispose a subject to markedly different types of actions. If I belief that I have been diagnosed with a deadly illness, I may be disposed write a will, contact loved ones, and shelve long-term plans. In contrast, the mere imagining of such a diagnosis is unlikely to have implications for motivation at all. Granted, if the episode of imagining is appropriately situated – say, as part of a therapeutic exercise in which I ask how I would live differently if I only had a short time to live – it may indeed influence my forward actions. But what is obvious is that the motivational profile of merely imagining the diagnosis will differ from that of a belief with the same content. The therapeutic exercise may convince me to reach out to my estranged relatives, but it will not motivate me to, say, liquidate my retirement plan.

The point is that imagining that *p* does not generally cause similar behaviors, or even similar kinds of behaviors, as believing that *p*.

Velleman recognizes that there are systematic differences between the motivational role played by beliefs and that played by imaginings. Typically when we imagine that *p* we also have a countervailing belief that not-*p*. If imaginings are to play an illuminating role in action explanation we need to know how belief exercises its priority over imagination. Velleman proposes that the relationship between beliefs and imaginings is one of "inhibition": pretend actions, at least in adults, are typically "inhibited" by countervailing real-life beliefs:

> Thus, for example, most deliberate imagining is accompanied by countervailing beliefs, embodying the subject's knowledge of the facts that he is imagining to be otherwise, such as his knowledge that an imagined pail of water is really a chair. These beliefs exert their worn motivational force, which can be expected to compete with that of the subject's imagination. Ordinary beliefs are not regularly accompanied by countervailing beliefs, and so their motivational force encounters less competition. I have also hypothesized that the motivational force of imagining comes under an inhibition, whose effects can be detected for example, in the way that we lower our voices when talking to ourselves. (272)

While Velleman's main point (that both pretense and belief have a role to play in rationalizing explanation) is persuasive, his picture of motivational forces that "outweigh" or "inhibit" each other is misleading. In a related footnote, Velleman describes the agent as acting on the "vector sum" of all motives combined (footnote 50). This locution, too, misses the mark. It is a mistake to think that the motivational force of the belief-like imagining and the desire-like imagining is "added up" to the countervailing force of belief and desire, and when the "vector sum" is positive one is moved to act, although in a qualified or mitigated way. To see why this picture misleads, consider the case *Sheila* adapted from D'Cruz (2008):

> Sheila is a calm and sympathetic person. But this morning she is irritated. Sheila has recently installed a new RAM module on her aging laptop to make it run more smoothly. To her great frustration, the computer runs even more slowly than it did before. Staring despairingly at the spinning rainbow wheel, she slams the lid of her laptop closed and mutters, "I give you a brand new hippocampus, and this is how you repay me? You *ungrateful* little shit." (D'Cruz 2008, 33)

It is obvious that if Sheila really believed that the computer was sentient and capable of understanding, she would not treat it as she does. Sheila's belief that the computer is *not* sentient and does *not* understand her muttering plays a key role in explaining her action. If Sheila had even the slightest doubt, she would never say something so callous. On the other hand, the cognitive attitude of imagining that the computer *can* understand her curses and *does* feel injured must also play a role in explaining Sheila's cursing at her

computer. Her action only makes sense in this light. Any explanation of Sheila's action that does not appeal to *both* the belief *and* the imagining would be insufficient to explain her action. Without the latter, her action would be inscrutable: why on earth does she scream at something that cannot hear, feel, or understand? Without the former, it would appear as if Sheila's behavior were wildly inconsistent with her character: how could Sheila abuse another so violently?

D'Cruz's case illustrates not only that both imagining and countervailing belief must play a role in the action explanation, but also that the relationship between belief and imagination is not one of simple "inhibition". Indeed, Sheila's countervailing belief that the laptop cannot feel or understand language explains why Sheila's action is so vehement. Were it not for this countervailing belief, she would not speak with such venom. To explain the action we need an account of the interaction between beliefs and imaginings. In addition, the relevant representations are not just the "countervailing beliefs" (i.e. – the computer cannot feel) but also beliefs about the context (e.g. no one else is listening).

In this spirit, Neil Van Leeuwen (2009) elaborates the systematic motivational differences between belief and imagining. The key difference is that imagining can only play a motivational role insofar as it is backed by a meta-representation, namely, a belief about the subject's "practical setting". Van Leeuwen draws inspiration for his account of the relationship between belief and imagination from an earlier account of relationship between belief and what Michael Bratman (1992) calls *acceptance in context*:

> An agent's beliefs provide the *default cognitive background* for further deliberation and planning […] This cognitive background is […] context-independent. But practical reasoning admits adjustments to this default cognitive background, adjustments in what one takes for granted in a specific practical context. […] To be accepted in a context is to be taken as given in the adjusted cognitive background of the context. (10-11)

Consider, for example, a risk-averse architect who makes cost estimations based on the upper end of projected range. Using Bratman's framework, we can say that the architect accepts that costs will be high in the context of risk even if she does not believe that costs will be high. Her belief that the costs will be moderate forms the *default cognitive background* for her *acceptance in context* that they will be high. It is on the basis of her beliefs about the practical setting she is in – e.g., presenting an estimate to a litigious client – that she determines to plan for the costs to be high. The choice to act on this adjusted cognitive background is typically automatic rather than consciously made. She simply plans for the costs to be high because of high stakes of underestimation. Her belief about the practical context forms the ground of her acceptance in context.

Van Leeuwen extends Bratman's framework to understand the relation between belief and imagining, arguing for what he calls the *practical ground thesis*: Beliefs are the practical ground of imagining (239). The basic idea is that non-belief attitudes only prompt action in virtue of being grounded in beliefs which determine if one is in the right practical setting for acting on the non-belief cognitive attitude. What distinguishes

imagining from belief is that the "psychomechanical efficacy" of imagining (the property of being effective in influencing action) is dependent on beliefs about the practical context. The agent must therefore represent the practical setting she is in to act with imagining as the adjusted cognitive background. Every time an agent performs an action on the basis of imagining, the enactment must be oriented by beliefs that represent the practical setting she is in. Beliefs determine if one is in the appropriate setting for guidance by imagination; imagination plays no such role for beliefs.

The practical ground thesis allows us to understand the intricate modulation by belief that we observe in pretend action. Consider, for example, the stage actor who, even while immersed in a role, takes measures to ensure that the theater audience can hear what she is saying. Consequently, she never faces away from the audience, and when she speaks she articulates more clearly and projects her voice further than she would in ordinary conversation. When she hears an audience member coughing, she adjusts her bearing so as to avoid distraction. But when she notices him getting red and gasping for air, she alerts the stage manager and stands to the side until medical help arrives. After the choking man is escorted away, she resumes from where she left off.[2]

The stage actor is a paradigm of what I will call a "witting pretender". Even while she is immersed in his performance, the knowledge that she is pretending shapes and constrains her performance. Not only can the actor skillfully modulate her manner of pretending, she can also direct rational scrutiny to the matter of when to break off from pretending and when to resume. Beliefs that represent the appropriate context for pretense are poised to be used in action-guidance as well as a premises in reasoning that generates further beliefs via inference.

Van Leeuwen (2011) points out that cognitive and conative attitudes "comment on and constrain imaginings" thereby influencing pretend action (67). In particular, such attitudes can "comment on the *value* to the agent of a particular imagined action." (67) We could add that such beliefs further "comment on" the comparative value of pretending in one particular way rather than another, and also on the comparative value of initiating or terminating pretense at one time rather than another. They allow the subject to fit pretense episode into her broader aims, and to execute the pretense in a way that satisfies rather than frustrates her real desires. Such beliefs also identify the etiology of representations, keeping track of whether they were formed via a process that is reality-sensitive (e.g. – visual perception) or via process that is reality-indifferent (e.g. daydreaming).[3]

In what follows I refer to this set of cognitive and conative representations collectively as the "comprehensive orientation representation". Since I will resort the phrase often, I give it an acronym: COR. Aspects of the COR may be explicitly articulated and scrutinized in anticipation of pretend action (for example, a group of live-action role players deliberating about the rules and boundaries of the game in light of their shared aims and

[2] This example was suggested to me by XXXXXXX.

[3] XXXX XXXX points out to me that there are also many interesting cases of pathologies where subjects lose track of the source (e.g. false recovered memory, paranormal experience). My characterization is only meant to describe typical cases.

desires). In more typical cases, the COR remains tacit unless major adjustments to pretend action are required (for example, the stage actor who needs to decide whether "the show must go on").

There is good reason to think that the COR is poised for action-guidance even with pretenders who are deeply immersed in an imaginative project (method actors, for instance). Liao & Doggett (2014) cite supporting evidence from developmental psychology indicating that children as young as three years keep track of the fact that they are pretending while engaged in immersive fantasy play even in the face of adult intervention designed to blur the boundary between fantasy and reality.[4] For instance, when an adult actually bites into a Playdough cookie (as opposed to merely pretending to bite), children are clearly shocked by the transgression ("Oh, you took a real bite. Now your teeth are all pink. How does it taste? … Yuck, do you always eat Playdough?") Here, the child's surprise indicates that despite her immersion in the imaginary game, she never loses track of the fact that she is only pretending that the "cookies" are edible. The authors of the original study conclude:

> On the basis of our findings it seems likely that the world of pretense and reality are not mutually exclusive. The playing child monitors events occurring in reality and maintains the duality between the two modes of thought. In this model of play, thoughts about reality run parallel to thoughts about pretense, although pretense would be the predominant mode of thinking during play. (Golomb & Kuersten 1996, 215)

This metacognitive "keeping track" accompanies even episodes of pretense that are never explicitly so framed, as is the case in spontaneous pretense. Even though Sheila never forms an explicit intention to pretend that her laptop is sentient, her actions are nonetheless guided and constrained by the COR. The accessibility of the COR explains why she is not surprised when the computer is not intimidated and why she does not later apologize for her shortness of temper. Hers is an ordinary case of expressive and imaginative action rather than an extraordinary case of temporary insanity. We find evidence of her sensitivity to practical context when, for example, she ceases muttering at her laptop when others enter the office.

Action-guidance by imagination does not typically result in practical irrationality when access to the COR is relatively global and relatively continuous. It is global when all aspects of the COR are accessible: what is imaginary and what is not; what the subject's aims are; what kind of pretend behavior might contribute to or threaten those aims. It is continuous when the subject has access to the COR at all moments during pretend action.

§2 Unwitting Pretense

The distinctive mark of "unwitting pretense" is that the accessibility of the COR for action-guidance is weakened. I concede that the very idea of unwitting pretense might

[4] The original study was carried out by Golomb & Kuersten (1996). Liao & Doggett cite a description of the study in Taylor (1999).

seem paradoxical. How could it be possible for a person to pretend without orientation? Wouldn't the person simply be performing real (non-pretend) actions? As we have seen, it is the continuous, global access to the COR that make possible the intricate modulation and circumscription of action that is characteristic of imaginative pretense. But I submit that this essential feature of a subject capable of imaginative pretense should not lead us to ignore the possibility that in some non-typical instances pretense can become episodically untethered from its ground in the COR. Extending a classic metaphor from Ramsey, Van Leeuwen proposes that beliefs are not only the maps by which we steer, but also the maps by which other maps are chosen and appraised (239). Extending this metaphor further, we might say that in cases of unwitting pretense, this latter "map of maps" is temporarily mislaid. What kinds of behaviors might such a state produce? Here is an illustration.

Snowpants. Consider the father of a recalcitrant three-year-old who refuses to put on his snowpants before going outside to play in the snow. In a gambit to turn conflict into cooperation, the father, instinctively and non-deliberatively, feigns outrage: "What do you mean, 'no snowpants'? Didn't we agree this morning that it's too cold and wet to go outside without them? You *promised* me that you would cooperate and wear your snowpants!" Of course, the father is not actually indignant; he is well aware that his young child has little or no grasp of "you promised". In the normal course of things he would try something else when these feigned expressions of indignation fail to gain traction. But this morning the father is tired, very tired. Grievances about other matters unrelated – the babysitter's last minute cancellation, the daycare tuition – lurk in the background. The father's counterfeit moral anger starts to feel like real grievance, and his ensuing words and actions betray this attitude. But before he moves to punish the child, he recalls the advice that he has so often dispensed: take a deep breath. It becomes vivid to him that his posture of indignation is merely tactical, that it is an "act". It becomes similarly vivid that the act is not effective in bringing about the wearing of snowpants. The father tries something different: distraction and bribery with chocolate.

In *Snowpants* preservation of the circumscription of settings in which the pretense-originated representation is allowed to guide action is tenuous. In the context of the disorientation brought on by strong emotion, the COR may become inaccessible for the purposes of inference and action-guidance. Such temporary loss and subsequent recovery of access constitutes irrationality of the garden-variety; it is not symptomatic of delusion or mental illness. It is a familiar fact about our mental lives that there are circumstances in which we can lose track of real and pretend; that is to say, there are periods when the COR is not continuously and globally accessible. Here is a another quotidian example:

Strategic Sulk: Leila had to stay late at work, leaving James to feed the kids and put them to bed on his own. Leila wasn't able to tell James that she would be late coming home until the last minute. Feeding and putting the kids to bed had not been overly difficult that evening, and there was nothing else James needed to be doing. But James has a manipulative streak. He perceives that there is leverage to be gained in playing the role of the aggrieved and he knows that Leila is vulnerable to guilty feelings. What's more, James has already committed to a weekend trip with some old college friends, and he has yet to tell Leila. James spends the next few days sulking and moping, looking for an

opportune moment to tell Leila about the trip. But James's victim act doesn't work. Rather than feeling culpable, Leila recognizes the guilt trip and responds with anger. But James's posture of being wounded has taken on a life of its own. He sabotages any attempt at reconciliation, making it less likely that he'll get the acquiescence he seeks.

Fervent yet phony emotional postures like the one described above are an ubiquitous predicament, and a perennial subject for writers of psychological realist novels.[5] A person who is pretending unwittingly behaves in ways that are characteristic of a person who labors under sincere but false belief. He may subvert his aims in ways that we would not expect of a witting pretender. Since the witting pretender has global and continuous access to the COR, he will generally pretend only in ways consistent with his desires, and he will stop pretending when guidance by imagination threatens to interfere with their fulfillment.

Having sketched both spontaneous and deliberate varieties of witting pretense, what of *unwitting* pretense? My hypothesis is that while there is logical space for deliberate unwitting pretense, it is not instantiated as psychological reality. To be more precise, while a subject who is engaging in deliberate imaginative pretense may lose access to the COR, it is not possible to purposefully initiate an episode of unwitting pretense. In consciously framing one's activity as guided by a representation that is reality-indifferent, it is not under a subject's control to render the COR inaccessible. (None of this is to deny that there may be indirect self-manipulations. It is sometimes possible to "fake it till you make it".)[6]

Examples similar to *Snowpants* and *Strategic Sulk* can be easily conjured once we start to notice them (contexts involving "open secrets" are a rich source). During such episodes, the subject does not abruptly *believe* something outlandish. The father in *Snowpants* does not, for any period, believe that his three-year-old is a moral reprobate for breaking his promise. Rather, it makes much more sense to interpret the father's behavior as that of a person who has temporarily lost track of the border between cognitive representations that are formed via a reality-sensitive processes and cognitive representations that are formed by processes that are reality-indifferent. His imagining inappropriately plays the action-guidance role usually reserved for belief. Similarly, James knows very well that Leila's lateness on a single occasion does not warrant guilt. His manipulative instinct to "play the victim" leads him to get carried away in a gambit for control, the sort of thing Leila could call him out on to force a reckoning.

---

[5] The phenomenon is also studied empirically by psychologists. Paul Griffiths and Andrea Scarantino (2009) suggest that sulking behaviors are modulated by strategic aims of which the subject is largely unaware at the time of their deployment. Although the sulking behavior is often in part performative pretense, subjects do not seem able to critically evaluate the aims of the performance contemporaneously. In a similar vein, expressions of anger, even if feigned, may bring about real feelings of anger. Cf. Bushman 2001.

[6] According to "status enhancement theory" people gain influence by acting dominant even if they do not feel that way on the inside. Expressing oneself with absolute conviction makes a person appear competent to others, which may in turn lead a person to lose track that she is pretending, and ultimately to believe that she is competent. Cf. Anderson et al 2012. Recent work by Luhrmann (2012) on the self-conscious religious pretending of American evangelicals suggests that it may even be possible to fake it *to* make it

It may seem tempting to think that episodes of unwitting pretense are mere marginalia in mental life. In the remainder of this paper, I make the case for the category's significance. I propose that appeal to the concept of "unwitting pretense" allows us to give the most perspicuous characterization of the state of mind of a person who is colloquially described as self-deceived.

§3. Self-Deception as Imaginative Pretense

The grouping of phenomena that gets the label "self-deception" in both colloquial and philosophical discourse is remarkably heterogeneous, ranging from processes of motivated irrationality to the state of full-blown delusion, from biased assessment of the evidence to outright denial. Because of this uneven terrain, any philosophical ambition to provide a unified account is bound to be quixotic.[7]

My target in this paper is the resultant cognitive state[8] of one particularly baffling sub-type of self-deception that I will refer to as "deeply conflicted self-deception".[9] I argue that this kind of self-deception should be modeled as a special variety of unwitting pretense.[10] The type of self-deception that is my explanandum is marked by three incongruous behavioral tendencies that I label "strategic evasion", "self-sabotage" and "surfacing":

> (1) *Strategic Evasion*: The subject avoids circumstances where he may be confronted with evidence that is inconsistent with his self-deceptive representations.
> (2) *Self-Sabotage*: The subject affirms and defends his self-deceptive commitment even when so doing undermines his goals.
> (3) *Surfacing:* In the context of a high-stakes forced choice, the subject abandons the self-deceptive commitment.

None of these behavioral tendencies is mysterious when considered in isolation. But the co-occurrence of these tendencies in the same subject with regard to the same body of evidence is deeply puzzling.[11] I will argue that the best way to make sense of them is to

---

[7] Egan (2008) and Van Leeuwen (2010) express similar skepticism.
[8] Along the way I will comment on the process of self-deception and the motivational structures that initiate it, but my primary explanandum is the resultant cognitive state.
[9] I borrow this phrase from Funkhouser (2009) although I use it somewhat differently. For Funkhouser what distinguishes this category of deeply conflicted self-deception is that there is no fact of the matter about what the subject believes. I reject this characterization although I am interested in the same set of cases.
[10] Following Liao & Gendler (2011), I use the term "imagination" to refer to our capacity simulate perspectives different from the one available through experience. This sense of (*recreative) imagination* is distinct from what Currie and Ravenscroft (2002) call *sensory imagination* (the willful capacity to have perception-like experience in the absence of relevant stimuli) and *creative imagination* (the capacity to combine ideas in unexpected and unconventional ways). "Imaginative pretense" or "pretense" refers to the guidance of behavior by imagination.
[11] Mele's influential deflationary account (2001) models self-deception as motivationally biased belief formation. Without denying that Mele's psychologically plausible account captures important and interesting facets of our mental life, the cases that comprise my explandum exhibit a specific kind of behavioral incongruity not easily accounted for with deflationary analysis.

understand the self-deceived subject as an unwitting pretender. Consider the case of *Alan*:

> Alan is self-deceived about the quality of this own work. Alan touts his brilliant scholarship to anyone who will listen, but it is apparent to everyone who reads it that the work is flimsy. Nonetheless, Alan badgers OUP to put out a volume of his collected papers. When he is confronted with his work's obvious shortcomings, he dismisses the criticism as jealousy and small-mindedness. He consistently ducks further discussion whenever there is a risk that his feeble arguments will be exposed to scrutiny. In particular, he comes up with ridiculous pretexts not to send his paper to Zeynep, his new junior colleague who knows the subfield well. Nonetheless, Alan is discerning when it comes to the work of other people; he makes accurate and uncontroversial judgments of better and worse. One of Alan's keenest desires is for acclaim in the profession. Unfortunately, his colleagues regard him with exasperation, pity, and bemusement. Those who care about him hope that he will someday emerge from this fantasyland.[12]

Confronting the full complexity of Alan's self-deceived state requires grappling with deeply incongruent behavioral tendencies. Alan's stubborn defense of the self-deceptive commitment (even to the point of losing all credibility and becoming the object of derision) appear to be evidence for his sincerity. But Alan's deft avoidance of contrary evidence suggests responsiveness on his part to the vulnerability of his attitudes to evidential override. Finally, we can imagine Alan jettisoning his self-deceptive commitment in a context of high-stakes forced choice. For example, with a promotion on the line, we wouldn't be surprised to find Alan soliciting Zeynep's help, leading us to think that he never really believed his work was flawless in the first place.

Deeply conflicted self-deceived subjects like Alan hold our interest because they appear at some moments oblivious and at some moments shrewd. They are strange to us, but also familiar. We can recall acting in similar ways in the past. This kind of epistemic tension led early theorists to posit homuncular and divided models of the mind to explain self-deception, notwithstanding all of the hopeless entanglements that such models invite.[13]

Tamar Gendler's model, SELF-DECEPTION AS PRETENSE, attempts to account for the full complexity of the most perplexing cases without paradox.[14] Drawing on work by David Velleman (2000), Gendler suggests that if we recognize the degree to which attitudes other than belief influence our emotions and govern our actions, the idea that self-

---

[12] The example is adapted from Doggett (2012, 765).

[13] For canonical elaborations of the homuncular and divided mind models, see Pears (1984) and Davidson (1985) respectively. Mark Johnston provides a good gloss on the perplexities such models invite: "How can the deceiving subsystem have the capacities to perpetrate the deception? […] Why should the deceiving subsystem be interested in the deception? Does it like lying for its own sake? Or does it suppose that it knows what it is best for the deceived system to believe?" (Johnston 1988, 64).

[14] Stephen Darwall (1988) also proposes a pretense account of self-deception, but I focus in this paper on Gendler's more recent and more detailed account.

deceivers are pretenders will not seem so strange. Gendler (2007, 2010) proposes that to be self-deceived is to engage in a form of imaginative pretense. On her view, Alan merely pretends that his work is brilliant, though he may know or strongly suspect that it is not. Alan's pretense comes to play a typically belief-like role in terms of introspective vivacity and action-guidance: in many contexts, Alan feels as if and acts as if he believes that his work is brilliant.

Gendler's model is designed to accord with the natural description that the phenomenon invites, of which both reality-sensitivity and reality-indifference are key features. For example, we might expect Alan to exhibit reality-sensitive tendencies when confronted with manifest and undeniable counter-evidence (say, a clearly articulated, knock-down objection from Zeynep), as well as the countervailing fantasy-maintenance tendencies of preventing himself from entering in to such circumstances (say, ignoring Zeynep's email offering to look at a draft of the manuscript). Moreover, Gendler thinks that modelling self-deceivers as pretenders comports well with the phenomenology of emerging from self-deception. The erstwhile self-deceiver does not feel as if he has given up one belief in favor of another with a contrary content; rather, he feels as if he has exited realm of fantasy and finally acknowledged something he either knew or suspected all along (2007, 245). The self-*deluded* subject who is debunked may express genuine surprise at what he has learned about the world; in contrast, the self-deceived subject who emerges from self-deception may be surprised by what he learns about himself.

Gendler proposes that the "cleanest and most interesting" cases of self-deception satisfy the following schema:

> (a) the person who is self-deceived about not-P pretends (in the sense of makes-believe or imagines or fantasizes) that not-P is the case, often while believing that P is the case and not believing that not-P is the case;
> (b) the pretense that not-P largely plays the role normally played by belief in terms of (i) introspective vivacity and (ii) motivation of action in a wide range of circumstances. (2007, 231)

Gendler's model accords with the inkling that "at some level" Alan knows or suspects that his work is flimsy (P). Nevertheless, Alan pretends (makes-believe) that the work is not flimsy but heterodox and brilliant (not-P), and the pretense (that not-P) plays the role normally played by a belief to that effect (he badgers OUP and shrugs off his friends' sincere advice). Alan's friends are likely to recognize in him the characteristic evasive maneuvers of someone in denial. In addition to evading Zeynep's critical eye, he may, for example, avoid presenting his work at colloquia where colleagues would press him on his insubstantial arguments.[15]

Gendler maintains that the self-deceiver's unusual relation to the evidence relevant to P is sustained by a certain kind of motivation: "the desire that, insofar as possible, she have the experience of being in a not-P rather than P world." (242) So motivated, self-

[15] Funkhouser (2005) usefully characterizes "avoidance behavior" as "avoiding evidence that not-p in a way that shows the agent already possesses sufficient information that not-p." (footnote 5).

deceivers mentally escape from the P environment and fashion for themselves a fantasy 'not-P-world'. This effectively renders them inhabitants of two disparate worlds, one in which they rationally believe P and another imaginary not-P-world that is buffered against hostile evidence.

This complicated relation to the self-deceptive content is revealed in three sorts of cases that Gendler dubs motivational occlusion, evidential override, and trumped incentive. In motivational occlusion, the motivation that perpetuates the self-deception is absent. For example, the disease sufferer who is self-deceived about the state of his health is able to make rational judgments when the identifying information is masked on his own medical records (243).

In trumped incentive cases – which Gendler thinks are the most compelling evidence for taking self-deception to be a form of pretense – some other goal matters more to the subject than the goal of maintaining the impression of being in a not-P world, and consequently he is willing to allow the rational belief that P to play its "rightful thought-occupying and action-guiding role." (244) For example, suppose a medicine becomes available that will confer dramatic benefits to the self-deceived disease sufferer. In such a context of high-stakes forced choice, the self-deceived subject will likely take the medicine, acting on the rationally based belief rather than on the imagining.[16] In contrast, if the subject is self-deluded rather than self-deceived, we should expect that he would decline the new medication.

Finally, "evidential override" occurs when the evidence for P is so overwhelming that to maintain not-P as the focus of thought and action would show a degree of reality-insensitivity that the subject finds intolerable. In such cases, the techniques of what I have been calling "strategic evasion" are not up to the task. Gendler's own examples of evasion include a subject who is self-deceived about her husband's affair and who is reluctant to drive past his purported lover's driveway for fear of seeing his car parked there[17], as well as a parent who is self-deceived about his child's innocence and who is reluctant to read newspaper reports for fear of coming across information linking her to the crime (244). The self-deceived subject avoids confrontation with these circumstances because they threaten to undermine the thought-occupying and action-guiding role the pretense plays in her life. In contrast, the *self-deluded* subject will take no such precautions. Gendler thinks that the ability to explain evidence avoidance and persistent resistance to evidential override is significant advantage for the pretense account: "Advocates of a belief account must explain why someone who believes that not-P—that is, someone who holds not- P with the aim of thereby holding something true—would so steadfastly avoid exposing herself to circumstances where P-relevant evidence might be available. The pretense account faces no such difficulty." (244)

But here we find an apparent disanalogy between self-deceivers and ordinary pretenders. Role-players and stage actors, for instance, may have no disposition and feel little pressure to avoid evidence that the content of what they imaginatively pretend is not real.

---

[16] Funkhouser (2005) makes the same point: "the behavioral dispositions of the self-deceived, especially when in situations where the costs of mistake are high, are tipped toward believing the truth." (307)
[17] This example is borrowed from Funkhouser 2005.

Moreover, a deceiver who pretends for the sake of misleading another person needs only to avoid the state of affairs in which the *other person* comes into contact with contrary evidence. But presumably the deceiver's own aims in pretense will not be subverted if the deceiver himself comes into contact with manifest counter-evidence.[18]

Nonetheless, I think that if we look closely, episodes of ordinary "pretense-for-play" and "pretense-for-deception" are not as different from Gendler's self-deceptive pretense as they may initially seem. Ordinary pretenders do take measures to sustain the *suspension of disbelief*, cultivating patterns of attention supportive of the kind of rich imaginative experience that not only produces behavior consistent with not-P but also engages emotion and desire that is consistent with not-P (which can in turn play a role in producing convincingly realistic not-P behavior). Children who are imaginatively immersed in a game in which they are superheroes will not direct their thoughts to the incongruous fact that they have recently completed their chores. Similarly, a confidence trickster or double agent may take measures to "stay in character" that include avoiding thoughts that are incompatible with her cover story. Although the subject's patterns of thought are supportive of suspension of disbelief, she never explicitly presents to herself that this is what she is up to. Doing so would be like trying very hard *not* to think of a pink elephant in the corner of the room. That is, it would have the ironic effect of upending suspension of disbelief.

Although there are important parallels between the subject who pretends in service of play or deception and the subject who self-deceptively pretends, there are also crucial differences. As I elaborate in the next section, the difference between the ordinary pretender and the self-deceiver is that the ordinary pretender does not lose track, not even episodically, of the fact that he guided in action by pretense rather than belief. Successful pretense-for-play or pretense-for-deception rests on a cognitive achievement that is conspicuously absent in self-deception.

§4. Self-Sabotage

While Gendler's SELF-DECEPTION AS PRETENSE promises solutions to some central puzzles of self-deception, it fails to meet a formidable challenge. Subjects who are self-deceived have a tendency to subvert their own goals, manifesting distinctive patterns of irrationality. But this is not typically true of participants in games of make-believe or of individuals pretending for the purposes of deceiving another person. Since SELF-DECEPTION AS PRETENSE models the self-deceived subject straightforwardly as a (witting) pretender, the self-sabotaging behavior characteristic of deeply conflicted self-deceivers remains deeply mysterious. In order to understand deeply conflicted self-deception in terms of imaginative pretense, we must be able to distinguish between the *witting* pretense of the actor or role player from the *unwitting* pretense of the self-deceiver. Only then will we arrive at an account that makes sense of why the self-deceiver at times appears cunning and at other times in the dark.

Consider first witting pretenders. Children will cease pretending to be cops and robbers

[18] I owe this point to XXXX XXXXXXXXXXX

when the game is no longer amusing, and sometimes when their parents tell them to stop. A confidence trickster will abandon his cover story when she learns that the game is up and it is counter-productive to try to sustain the subterfuge. But in marked contrast, self-deceivers will often maintain and defend their self-deceptive commitment even when so doing undermines their most cherished goals. Alan, for example, may maintain his postures of brilliance even when they consistently result in ridicule and scorn from his colleagues whose admiration he craves. Alan's self-sabotage is symptomatic of fact that, in contrast with ordinary pretense, self-deceivers exhibit distinctive patterns of irrationality. This is a crucial explanandum for any satisfactory model of deeply conflicted self-deception.

Tyler Doggett (2012) notices that distinguishing between ordinary pretense and self-deception is a key element missing from Gendler's SELF-DECEPTION AS PRETENSE, which he describes as an otherwise "extremely intriguing and promising view" (765). He suggests that, absent elaboration, the pretense account does not accord with our sense that self-deception is an irrational condition. After all, there is nothing particularly irrational about pretending that *P* while knowing that not-*P*. Doggett describes the case of Bob, aged 6 years, obsessed with trolls, and deeply immersed in an imaginative project of acting and emoting as (he imagines) a troll would. Doggett notes that, "Whatever his failings, Bob isn't irrational. Annoying. Obstinate. Childish. Not irrational." (766) The problem for SELF-DECEPTION AS PRETENSE is that both Bob and Alan appear to satisfy Gendler's characterization of the self-deceived subject. Each "pretends that [P] is the case, […] while believing that [not-P] is the case and not believing that [P] is the case" and the "pretense that [P] largely plays the role normally played by belief in terms of (i) introspective vivacity and (ii) motivation of action in a wide range of circumstances." (Gendler 2007, 231)

We may hope to find the distinctive mark of self-deceptive pretense in its etiology. Perhaps self-deceived subjects, unlike actors, pretend that *P* because they wish that *P* were true. But this will not mark off the two kinds of pretense correctly. Bob's motivation for pretending to be a troll could easily be his keen desire to be a troll. But this would not make it the case that Bob is self-deceived. In the case of Bob, the pretense that *P* plays the role normally played by belief. In the case of Alan, the pretense that *P* *inappropriately* plays the role normally played by belief. The challenge for the exponent of SELF-DECEPTION AS PRETENSE is to explain this discrepancy. Only then will the account "accord well with the sense that self-deception is an irrational condition." (Gendler 159)

Christoph Michel and Albert Newen raise a related concern for SELF-DECEPTION AS PRETENSE. They describe Gendler's concept of 'pretense' as "a hybrid that eats its cake and keeps it, too". They charge that Gendler's picture of the pretense implicated in self-deception is "belief-like in explaining S's [not-P] behavior while being sufficiently imagination-like so as not to conflict with S's knowledge that [not-P] is untrue." (737) They point out that if Gendler's pretense is functionally similar ordinary pretending, "one will not be able to explain how self-deceivers establish and defend [not-P] against acute evidence and challenges by others." (737)

A third critic of Gendler's SELF-DECEPTION AS PRETENSE model, José Eduardo Porcher,

offers two cases that emphasize the self-deceiver's tendency toward self-sabotage. First, Porcher considers an indebted businessman who over-estimates his own skills and ignores the failure of his past and current ventures: "he is not just stubborn, but adamant, and won't listen to reason, won't extract from the evidence that same conclusion anyone else would, and won't heed the advice of friends and family." (2014, 323) Porcher emphasizes the fact that subjects like this often end up subverting their own cherished goals: "It is not in the best interest of people like the businessman to perpetuate and plunge even further into his already outstanding debt. But he deliberately does just that." (323) Porcher's second case involves a single mother who welcomes a new boyfriend into her home despite the heaping evidence that he is sexually interested in her daughter. The mother who "is blindly in love with the man" refrains from asking any probing questions (324). As a result, she exposes her daughter, whom she loves deeply, to the risk of abuse.

Porcher expresses skepticism that serious and potentially hazardous courses of action such as these could be undertaken on the basis of mere imaginative pretense.[19] He maintains that if imaginings are psychomechanically effective in cases of self-deception, then the agent must have beliefs that represent this "peculiar practical setting." (322) The agent must therefore believe that she is pretending. But Porcher thinks that "it is plainly impossible to be motivated to act on our self-deception if not only do we not believe their content, but we simultaneously believe we are only pretending that such content is true." (322) Gendler's solution is thus vulnerable to a species of the dynamic puzzle of self-deception: "in order to voluntarily act in the context of imaginative pretense, the self-deceiver would have to know that she is in the context of imaginative pretense and choose to act on the basis of such pretense – and this, in turn would make for a self-refuting project." (327)

The specific objections from Doggett, Michel & Newen, and Porcher all issue from the same fundamental worry. The very possibility of acting on imaginative pretense seems to require that pretenders must represent the practical setting as appropriate for action-guidance by imagination. But this makes it is unaccountably mysterious that self-deceivers defend their self-deceptive commitments to the point of undermining their most cherished goals. This tendency to self-sabotage seems to suggest that the self-deceptive commitment is a sincere but false belief.

In "The Product of Self-Deception" (2007) Neil Van Leeuwen adduces two further reasons to think that the product (i.e. the resultant state) of self-deception is a belief.[20] First, he argues that the product of self-deception governs other cognitive attitudes, such as hypothesizing that P, in the manner characteristic of belief. His example is that of a

[19] Porcher notes that Van Leeuwen (2007) mounts a similar objection against the avowal accounts of self-deception of Audi (1982) and Rey (1988). I discuss Van Leeuwen's view in the section that follows.
[20] Van Leeuwen's explicit target in this paper is the avowal account of self-deception due to Robert Audi and (1982) and Georges Rey (1988). According to the avowal view of self-deception, the subject who is self-deceived that P is disposed to avow P, but does not believe that P. (Avowals are, roughly, dispositions to affirm propositions that lack deep connections to action). Van Leeuwen objects to this view, arguing that the product (i.e. the resultant state) of self-deception must be a belief.

father who is self-deceived about his son's intelligence. Van Leeuwen suggests that the father will likely hypothesize that his son's future grades will be high, and if they turn out to poor, he will likely reject the hypothesis that the grades are reflective of his intelligence. Van Leeuwen takes this to be evidence that the self-deceptive attitude "governs" the attitude of hypothesis insofar as it plays a systematic role in determining its contents (435).

Van Leeuwen's second reason to think that the product of self-deception is belief is that its "psychomechanical efficiency" (its effectiveness in influencing action) is comparatively strong in comparison to other cognitive attitudes such as supposition. Cases like that of the self-deceived businesswoman who takes out one loan after another to keep her failing business afloat show that the self-deceptive cognitive attitude is psychomechanically effective in a wide variety of contexts: "It is likely that its influence on action pervades many contexts such as speech, planning various endeavors, taking out loans, and considering whether to quit. If this is the case, then the product of self-deception forms the default for action relative to other, context-sensitive attitudes." (435-436)

Doggett, Michel & Newen, and Porcher are surely right that self-undermining adherence to the self-deceptive commitment is an essential feature of the explanandum. And Van Leeuwen is surely right that the resultant cognitive state of self-deception influences action across a broad range of contexts. But these insights are fully consistent with modeling the self-deceiver as an *unwitting* pretender. While the phenomenon of self-sabotage is a strong objection to Gender's original formulation of SELF-DECEPTION AS PRETENSE, it is naturally accounted for when we notice that pretense can be unwitting.

Recall the subject who is pretending unwittingly behaves in ways that are characteristic of a person who labors under sincere but false belief. He may subvert his aims in ways that we would not expect of a witting pretender because he lacks global and continuous access to the COR. If self-sabotaging behavior is consistent with postulating that the self-deceptive commitment is an imagining, and also consistent with postulating that the self-deceptive commitment is an belief, then we need a further reason to choose one account over the other. We find that further reason by paying attention to the emergence from self-deception.

§5. Surfacing

If we look more closely we will see that the product of self-deception does not influence behavior in the way that we would expect of a belief with the same content. To see this, think back to the phenomenon of "trumped incentive" highlighted by Gendler. Recall that in these circumstances, some other goal comes to matter more to the subject than the goal of maintaining the impression of being in a "not-P world", and consequently the rational belief that P is allowed to play its "rightful thought-occupying and action-guiding role." (244) This brings the erstwhile self-deceiver to abandon his self-deceptive commitment. For an illustration we can elaborate on Van Leeuwen's own examples.

Consider first the case of the self-deceived businesswoman who falls deeper and deeper

into debt, insisting to her exasperated family and friends that rich profits are just over the horizon. Suppose that, unexpectedly, a wealthy investor who wants to control the building in which her business is housed offers to pay her a hefty sum for the business, much more than it is really worth (which is almost nothing). In response to this unlikely opportunity, she cuts her losses and sells up, seizing her big chance to escape the pervasive anxiety of running a failing enterprise. Consider next Van Leeuwen's case of the father who is self-deceived about the academic ability of his son. Suppose that at a critical stage in his schooling, the son has an opportunity for free access to an otherwise unaffordable but highly regarded remedial academic service. It would not surprise us that with so much on the line, the father seizes the opportunity thereby acknowledging the fact that his son is in need of academic help. Van Leeuwen is right that the "mere avowal" account makes the self-deceiver's self-undermining behaviors unaccountably mysterious. But the supposition that the self-deceptive commitment is a belief makes the phenomenon of surfacing from self-deception in the context of high-stakes forced choice equally mysterious.

It must be granted that in either of these examples we can just as easily imagine the self-deceived agent stubbornly defending his self-deceived commitment, exposing himself and his loved ones to greater and greater hazard. Indeed, self-sabotage is one of characteristic marks of the kind of self-deception that is my explanandum. But what makes the central cases of self-deception of such enduring philosophical interest is that they are, as Gendler puts it, "both strangely precarious and strangely resilient." (2007, 245) The resilience of the self-deceptive commitment is observed in the self-sabotaging tendencies, typically indicative of (sincere but false) belief. The subject's awareness of the precariousness of the self-deceptive commitment is observed in evidence-avoidance tendencies and the surfacing tendencies, typically suggestive of guidance by reality-sensitive representations. The strangeness comes from the difficulty of interpreting these contradictory tendencies in the same subject with regard to the same topic.

One might worry that if the COR is sufficiently accessible to guide strategic evasion of evidence, then it becomes mysterious how self-deception could be sustained. Clearly, the self-deceiver never explicitly weighs the costs and benefits of an imaginative project aimed at experiencing life as if not-P against the costs and benefits of allowing himself to be guided by the rational belief that P. Were she to do this, she would not be self-deceived. But we need not suppose that the guidance by the COR observed in episodes of strategic evasion requires *paying attention* to the COR. This is because guidance by and awareness of something does not typically require paying attention to it. Consider this illustration taken from Arpaly & Schroeder (2014): Emma is a philosopher walking through a park and thinking hard about the argument of a paper she is writing. Completely focused on how her argument will go, she walks around a tree that was in her path. She was not ignorant of the tree. After all, she managed to avoid it. At the same time, she did not pay attention to it either; her attention was narrowly focused on the paper she was writing.

In avoiding circumstances that threaten exposure to evidence that is incompatible with the self-deceptive commitment, the self-deceiver is guided by a representation of own mind in much the same way that Emma is guided by a representation of the tree. In

particular, the self-deceiver is guided by a representation that her self-deceptive commitment, having been formed via process that is reality-insensitive, is vulnerable to evidential override. This cognitive representation may guide fine adjustments to action without being the focus of attention.[21]

Maiya Jordan (2017) proposes, plausibly I think, that the phenomenology of deeply conflicted self-deception typically includes a felt disquiet: self-deceivers are uneasy with their self-deceptive commitment, betraying an awareness that the truth is "dangerously close at hand." (2) Their commitment is "characteristically fragile, and so experienced." (2) It is not unusual for features of our phenomenal awareness to cause uneasiness even when our attention is not focused on them. Consider the uneasy mood created by a loud ticking clock or the sound of barking dogs: one cannot always pinpoint the source of the anxiety or even the fact that one is anxious.

A subject's patterns of attention and her desires are closely linked.[22] A desire for the success of a cherished project that is suddenly imperiled, or a desire for the flourishing of a loved one who is dramatically endangered, will have a tendency to shift attention to these matters. It is not surprising to observe shifts in attention in contexts where a subject is guided by reality-insensitive representations in a practical setting where such guidance dramatically threatens something the subject cares about. This is why a person's patterns of emergence from self-deception can reveal aspects of what that person desires, values, or cares about. Consider once again Porcher's example of the self-deceived single mother who refuses to face up to the growing body of evidence that her boyfriend is sexually interested in her daughter. The mother displays the typical avoidance tendencies of the self-deceived when it comes to evidence of the boyfriend's obsession and leering attention. But now imagine that she notices that her young daughter has begun to behave in ways that express indignation and that suggest injury. This, it turns out, is too much for the mother to ignore.[23] The salient desire for her daughter's flourishing is a part of the explanation for why she emerges from self-deception when she does. It becomes impossible to ignore that her commitment to her boyfriend's innocence is nothing but a self-insulating fantasy.

Having recognized one's self-deceptive postures as such, it is impossible to re-inhabit the point of view of being self-deceived right away. The fantasyland of self-deception, once

---

[21] Nisbett & Wilson's (1977) confabulation studies provide empirical evidence of complex reasoning that is unavailable for report. Also relevant is Armstrong's (1980) well-known example of "coming to" at the wheel on a long drive. Behavioral evidence demonstrates that the driver was guided by a representation of the road before the "came to". In this sense, the driver was "aware of" or "conscious of" the road. But the subject was not aware of this awareness. So in a different sense, the decisions made as a driver were unconscious.

[22] This is pointed out well by Thomas Scanlon (1998, 39ff.) and Nomy Arpaly (2011, 77). The famous "cocktail effect" of hearing one's name at crowded party illustrates that involuntary shifts in attention can be controlled by processing that is unconscious. Cf. Moray (1959)

[23] In Porcher's version of the story, witnessing her daughter's distress does not upend the mother's self-deception. Given different patterns of psychological salience and care, this, too, is a plausible version of the story. Herbert Fingarette (1969) argues, persuasively, I think, that overcoming self-deception consists in learning how, through a combination of therapy and "courage", to "spell out" or make explicit to oneself what one is really up to (130).

exited via reckoning, is not available for quick re-entry.[24] This feature of the self-deceptive state of mind often goes unnoticed, and what it reveals about self-deception is rarely appreciated.[25] Emergence from self-deception via reckoning means that the relevant COR becomes globally accessible, and it is not within a person's direct voluntary control to render it inaccessible. None of this applies to the (witting) imaginative pretense of the actor who can always take a deep breath and resume pretending.[26]

The picture of deeply conflicted self-deception that emerges of one of a syndrome composed of diachronic episode types. During episodes of strategic evasion the COR is accessible for action-guidance, although it never becomes the object of awareness. During episodes of self-sabotage, access to the COR is occluded. The self-deceiver loses track of the appropriate practical setting for guidance by reality-insensitive representations, and may therefore end up subverting her own goals. Finally, reckoning with the fact that one is self-deceived affords global and continuous access to the COR and is therefore constitutive of surfacing from self-deception.[27]

§6. Conclusion

The idea of unwitting pretense becomes less paradoxical when we see it as a limiting case of typical pretend action. Noticing that pretend action can become episodically untethered from its grounding in the COR allows us to make sense of behaviors that would otherwise be inscrutable. In particular, modeling self-deception in terms of unwitting pretense accords with our understanding of self-deceivers as both tactically evasive and self-sabotaging, as well as with our understanding of self-deception as resilient in some circumstances and unstable in others.

One of the earliest explorations of self-deception in the Western canon comes to us from Second Samuel. King David, drunk on lust and on power, has an affair with the ravishing Bathsheba, wife of the righteous soldier, Uriah. In a scheme to get rid of Uriah, David sends him to the fiercest part of the battlefront. When he gets word of Uriah's death, David allows Bathsheba to mourn for a short period, and then marries and impregnates her. The Lord is not content to allow David to inhabit the fantasy that he is exempt from the law, and sends Nathan to disabuse him of the conceit. Rather than confronting David directly, Nathan tells David a story about a rich man with many animals and a poor man with just one ewe lamb that "did eat of his own meat, and drank of his own cup, and lay

[24] Chance et al. (2015) report the finding that self-deception is "quickly revived" in lab settings. However, the subjects in the study never *reckon* with having been self-deceived. Rather, they display persistent over-prediction of performance even after observing themselves perform poorly.

[25] Exceptions include L. Jonathan Cohen who remarks, "once you accept that you have spotted self-deceit in yourself on some issue, it has presumably thereby ceased to exist in you on that issue" (1992, 147) and Funkhouser (2005) who refers to the fact that the self-deceiver cannot believe that she is self-deceived as the *opacity of self-deception*.

[26] XXXX XXXX pointed out to me that an actor could have difficulty regaining her composure and recovering after an interruption. But this would not be due to structural features of witting pretense as such, but to individual psychological difficulties.

[27] This is not to deny that there are cases where a person ceases to be self-deceived without ever undergoing such a reckoning.

in his bosom, and was unto him as a daughter."[28] When a traveler comes to visit the rich man, he takes the cherished lamb of the poor man and offers it to the traveler instead of parting with his own riches. David, incensed, demands that the rich man be punished by death. Nathan's notorious reply to David: "You are the man."

Nathan's allegorical tale allows David to recognize that he inhabits an unreal kingdom of his own creation, in which he is a kind of god. This is a reckoning for David rather than an epiphany. By acknowledging the self-deceptive fantasy as such, David is expelled from this realm, and there is no easy way to get back in. In Gendler's terms, David's recognition that he is guided by fantasy rather than belief is made possible by "motivational occlusion". David makes a judgment about a case that is outside the ambit of his insulating fantasy, and then is brought to see a parallel between that case and his own situation. The story reveals the important truth that often we lack the strength of conscience and the moral courage to see our self-deceptive fantasies for what they are, which is what is required if we are to surface from and reckon with self-deception. A fruitful avenue of future ethical inquiry is the role of honest and perceptive friends in playing Nathan to our David, that is, in getting us to notice that we are pretending.

[28] 2 Samuel 12, King James Version (KJV).

References

Audi. R. 1982. Self-Deception, Action, and Will. *Erkenntnis* 18:133-58.

Anderson, C; Brion, S; Moore, D.; Kennedy, J A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*. 103(4) 718-735.

Armstrong, D. M. 1980. *The Nature of Mind and Other Essays*. Queensland: University of Queensland Press.

Arpaly, N and Schroeder, T. 2014. *In Praise of Desire*. Oxford University Press.

Bratman, M. E. 1992. Practical Reasoning and Acceptance in a Context. *Mind* 101: 1-15.

Bushman, B. 2002. Does Venting Anger Feed or Extinguish the Flame? Catharsis, Rumination, Distraction, Anger, and Aggressive Responding. *Personality and Psychology Bulletin* 28:6, 724-731.

Chance, Z.; Gino F.; Norton M.; Ariely, D. 2015. The slow decay and quick revival of self-deception. *Frontiers in Psychology*. 6: 1075

Cohen, L. J. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.

Currie G, Ravenscroft I. 2002. *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.

Darwall, S. (1988). Self-deception, autonomy, and moral constitution. In B. Mclaughlin & A. O. Rorty (Eds.), *Perspectives on self-deception* (p. 407- 430). University of California Press.

Davidson, D. 1985. Deception and Division. In : J. Elster (ed) *The Multiple Self*. Cambridge University Press.

D'Cruz, J. 2008. Action and Imagination in Action. (diss.) Brown University electronic repository.

Doggett, T. 2012. Some Questions for Tamar Gendler. *Analysis* 72: 764-774.

Egan A. 2008. Imagination, delusion, and self-deception. In: Bayne T, Fernandez J, eds. *Delusions, Self-Deception, and Affective Influences on Belief-formation*. New York: Psychology Press.

Fingerette, H. 1969. *Self-Deception*. London: Routledge and Kegan Paul.

Funkhouser, E. 2005. Do the self-deceived get what they want? *Pacific Philosophical Quarterly* 86: 295–312.

Funkhouser, E. 2009. Self-Deception and the Limits of Folk Psychology. *Social Theory and Practice* 35(1): 1-13

Gendler, T. 2007. Self-Deception as Pretense. *Philosophical Perspectives*, 21: 231-258.

Golomb, C. and Kuersten, R. 1996. On the transition from pretence play to reality: What are the rules of the game? *British Journal of Developmental Psychology* 14: 203–217.

Griffiths, P. & Scarantino, A. 2009. Emotions in the Wild: The Situated Perspective on Emotion, In P. Robbins and M. Aydede (Eds.), *The Cambridge Handbook of Situated Emotion* (pp. 437-453) New York: Cambridge University Press.

Holton, R. 2000/1. "What is the role of the self in self-deception?" *Proceedings of the Aristotelian Society* 101: 53–69

Hume, D. 1776/1977. *An Enquiry Concerning Human Understanding*. Hackett Publishing Company.

Hursthouse, R. 1991. Arational actions. *Journal of Philosophy* 88 (2):57-68.

Jarrold, C., Carruthers, P., Smith, P. K. and Boucher, J. 1994. Pretend Play: Is It Metarepresentational? *Mind & Language* 9: 445–468.

Jordan, M. (MS). Self-Deception and Intentionality: On the Shared Error of Intentionalism and Deflationism

Johnston, Mark (1988). "Self-deception and the nature of mind." In . In: B. McLaughlin and A. Rorty, (eds.) *Perspectives on Self-Deception*. University of California Press.

Leslie, A. M. 1987. Pretence and representation: The origins of 'theory of mind'. *Psychological Review*, 94, 412-426.

Liao, S. and T. Doggett. 2014. The Imagination Box. *Journal of Philosophy* 111 (5):259-275.

Liao, S. and T. Gendler. 2011. Pretense and Imagination. *Wiley Interdisciplinary Reviews* 2 (1):79-94.

Mele, A. 2001. *Self-Deception Unmasked* (Princeton, NJ: Princeton University Press).

Michel, C. and Newen, A (2010). Self-deception as pseudo-rational regulation of belief. *Consciousness and Cognition* 19 (3):731-74

Moray, N. (1959). "Attention in dichotic listening: Affective cues and the influence of instructions" *Quarterly Journal of Experimental Psychology*. **11** (1): 56–60.

Nisbett, R.E. and T. Wilson. 1977. On saying more that we can know: Verbal reports on mental processes. *Psychological Review* 84:231-259.

Pears, D. 1984. *Motivated Irrationality*. Oxford University Press.

Porcher, J. E. 2014. Is Self-Deception Pretense. *Manuscrito* 37:3, 291-332.

Rey, G. 1988 Toward a computational account of akrasia and self-deception. In: B. McLaughlin and A. Rorty, (eds.) *Perspectives on Self-Deception*. University of California Press.

Rosenthal. D. 2000. "Consciousness and Metacognition", in *Metarepresentation: Proceedings of the Tenth Vancouver Cognitive Science Conference*, ed. Daniel Sperber, New York: Oxford University Press, pp. 265-295.

Scott-Kakures, D. 2002. "At 'permanent risk': Reasoning and self-knowledge in self-deception." *Philosophy and Phenomenological Research* 65: 576–603.

Taylor, Marjorie. 1999. *Imaginary Companions and the Children Who Create Them*. New York: Oxford University Press.

Van Leeuwen, N. 2007. The Product of Self-Deception *Erkenntnis* 67: 3, 419-437.

Van Leeuwen, N. 2009. The Motivational Role of Belief *Philosophical Papers* 38:2, 219-46.

Van Leeuwen, N. 2010. Why Self-Deception Research Hasn't Made Much Progress http://theblog.philosophytalk.org/2010/09/why-self-deception-research-hasnt-made-much-progress-.html Retrieved Aug 21, 2014.

Velleman, J. D. 2000. The Aim of Belief. In *The Possiblity of Practical Reason*. Oxford University Press.