Self-deception and the second factor: How desire causes delusion in anorexia nervosa

Final Draft, Do Not Cite

Forthcoming in *Erkenntnis*

Stephen Gadsby
Monash, Melbourne VIC, Australia
Gadsby.st@gmail.com
ORCID: 0000-0002-2302-844X

Acknowledgements

Thanks to Peter Clutton & Jakob Hohwy for feedback on earlier drafts. This research was supported by an Australian Government Research Training Program (RTP) Scholarship.
Abstract

Empiricist models explain delusional beliefs by identifying the abnormal experiences which ground them. Recently, this strategy has been adopted to explain the false body size beliefs of anorexia nervosa patients. As such, a number of abnormal experiences of body size which patients suffer from have been identified. These oversized experiences convey false information regarding the patients’ own bodies, indicating that they are larger than reality.

However, in addition to these oversized experiences, patients are also exposed to significant evidence suggesting their bodies are in fact thin. This situation poses a conundrum: why do patients appear strongly influenced by the former kinds of evidence while the latter has little effect?

To solve this conundrum, I suggest a two-factor account. First, I discuss research on the biases patients exhibit in how they gather, attend to and interpret evidence related to their own body size. Such biases in evidence treatment, I suggest, cause oversized experiences to be sought out, attended to and accepted, while veridical body size experiences are ignored or explained away. These biases constitute the second factor for this empiricist model, accounting for the unwarranted conviction with which these beliefs are held. Finally, in line with recent research into self-deception, I propose that, paradoxically, these biases in evidence treatment arise from patients’ own desires.
1. Introduction

Many clinical patients assert the truth of patently implausible statements: some claim that a loved one has been replaced by an imposter (Capgras delusion), others, that a part of their body belongs to someone else (Somatoparaphrenia) and some even insist that (despite appearances) they are, in fact, dead (Cotard delusion). These delusions can arise in the context of a diverse number of conditions, such as brain trauma, cerebral accident, degenerative brain conditions and substance use disorders. Overlooking this bewildering diversity, there is an incredibly fertile field of cognitive neuropsychiatry devoted to understanding the nature and cause of the beliefs themselves (Coltheart 2007).

In contrast with these patently bizarre delusions, the false beliefs I will be discussing here pertain to the somewhat mundane topic of body size and are generally found in a small subset of related eating disorders. Eating disorders (ED) are characterised by persistent disturbed eating related behaviour of some form—an underspecified characterisation which aggregates a number of disorders with distinct and diverse symptomatology, and likely aetiology.

Perhaps due to the emphasis on eating behaviour, the false body size beliefs of ED patients have avoided the level of scrutiny applied to the aforementioned delusions. Indeed, psychiatrists have traditionally refused adopting the label of delusion to describe these beliefs, instead classifying them as overvalued ideas—a form of belief which carries an “air of limited clinical significance”, and is regularly associated with abnormal personality, rather than pathology per se (McKenna 1984, 579; see also: Veale 2002). In spite of the orthodox line, a move is now occurring towards classifying at least a sub-group of these beliefs as delusions; with this, comes increased scientific interest into their nature and cause.
While some of the evidence I will discuss pertains to ED patients more generally, the lion’s share pertains to anorexia nervosa (AN) specifically. For the time being then, I restrict my claims to these specific patients. Nevertheless, I remain open to the possibility that the following model might be adapted to account for the abnormal body size beliefs of others, such as bulimia nervosa (BN) patients (see section 4). Despite undeniable differences in symptomatology, the various eating disorders may share important commonalities in the aetiology of the patients’ harmful body size beliefs.

In order to pin down the relevant target of scientific interest, a first task is to describe the content of such beliefs. At first glance, this appears problematic, as specific belief content undoubtedly varies from patient to patient. Despite common misconception, not all AN patients believe they are “fat” so much as that they aren’t “thin”, or even that they aren’t “excessively/dangerously/too thin” (Konstantakopoulos et al. 2012; O’Connell et al. 2017). The inherent issue is that specific belief content is largely determined by idiosyncratic standards regarding concepts such as “thinness” and “fatness”.

That said, this diversity in belief content need not concern us. In order to pin down a specific framing of belief content shared by AN patients, I focus here on beliefs regarding not yet meeting one’s ideal body size. This highlights the fundamental irrationality of patients’ beliefs, as the central problem is not that they have unrealistic desires or extreme evaluative standards, but rather that they do not realise they have already met and surpassed their own ideal standards for thinness (Moscone et al. 2017). Further, while specific belief content—e.g. “I am not thin”, “I am fat”, “I look awful” etc.—might vary, most patients can be said to believe they aren’t yet their ideal size (Gardner & Brown, 2010).
2. The empiricist approach to anorexia nervosa

One approach which has seen considerable success in explaining delusions is termed *empiricism* (Bayne & Pacherie 2004). Emerging from the work of Maher (1974), empiricist models aim to explain delusional beliefs by identifying abnormal experiences with related content which might ground them. To take an example from the aforementioned delusions, in the standard empiricist model of Capgras it’s proposed that a breakdown in the system responsible for autonomic responses to familiar faces causes these patients to lack the usual sense of familiarity when seeing a loved one (Ellis & Young 1990). This abnormal experience (recognising a loved one but lacking the expected affective response) is assumed to ground the belief that they are an imposter.

Adapting this explanatory strategy to the case of AN has also proven fruitful as, on inspection, there are a number of abnormal experiences of body size which patients suffer from that may ground their abnormal body size beliefs—what have been termed *oversized experiences* (Gadsby 2017a; 2017b). It should be noted that these oversized experiences are largely visual in nature. While there is a significant body of evidence tying eating disorders to dysfunctional interoceptive processing (Kaye et al. 2009; Pollatos et al 2008), it seems unlikely that one could experience their bodies as larger than reality through interoception alone—as interoceptive signals do not deliver content related to body size or shape.\(^3\) In what follows, I briefly review this empiricist model and the oversized experiences it posits.

2.1. Oversized Experiences

One form of oversized experience AN patients likely suffer from is *spontaneous recurrent mental imagery*. Most commonly seen in anxiety disorders, this involves constant, intrusive
mental imagery representing an individual’s own fears (Hackmann et al. 1998). There is some evidence to suggest that ED patients suffer from recurrent mental imagery, whereby the images are of their own bodies as overweight (Cooper et al. 1998; Somerville & Cooper 2007; Somerville et al. 2007). It’s possible that such recurrent imagery could ground false beliefs about body size, as long as patients took this imagery to be veridical. While those who suffer from anxiety disorders appear to interpret their imagery as veridical (Hackmann et al. 2000, 602), it’s not yet clear whether this also holds for ED patients.

Another form of oversized experience stems from perceptual body image distortion. Following Schilder (1935, 11), the perceptual body image can be defined as “the picture of our own body which we form in our mind”. In contrast to spontaneous episodes of mental imagery, the content of the perceptual body image derives from a stored representation we have of the shape, size and form of our own bodies (de Vignemont 2010; Gadsby 2017c). In the case of AN, this mental image is said to be distorted, as patients’ bodies are represented as larger than reality. Evidence of this comes from tasks requiring patients to estimate the size of their bodies, whereby overestimation is considered reflective of an oversized perceptual body image (Smeets 1997).

How could distortion of the perceptual body image cause oversized experiences? One possibility is through a mode of experience similar to the previously discussed episodes of spontaneous mental imagery. AN patients may consciously mentally picture this distorted body representation and, similar to spontaneous mental episodes, take it to be veridical. Although this story seems plausible, there is, as of yet, no evidence of such behaviour.

Nevertheless, there is a relevant kind of activity ED patients regularly engage in: mentally comparing themselves with their “thin” peers (Alleva et al. 2013, 99; Corning et al. 2006; Espeset et al. 2012, 524; Hamel et al. 2012). This process of size comparison relies on the perceptual body image, as judging whether a visually presented body is larger or smaller
(without looking at our own bodies) requires mental comparison with one’s own perceptual body image (Longo 2015; Longo & Haggard 2012). In fact, this process is markedly similar to that of many body size estimate tasks, which require patients to compare their own body size to a number of visually presented bodies (Gardner & Brown 2010). Patients’ misjudgement in these tasks suggests similar misjudgement may occur throughout regular life: when comparing their body size with others, patients would falsely judge themselves to be larger. These regularly occurring (mis)judgments might plausibly ground beliefs about not yet reaching one’s ideal size.

Finally, it has been proposed that patients falsely misjudge the *affordances* of their environments. Affordances can broadly be defined as the actions provided by an agent’s environment (Gibson 1979). In the case of AN, false affordance judgment arises from distortion of the *body schema*—a sub-personal representation of the body relied on for motor control and affordance processing (de Vignemont 2010; Gadsby 2017c; Keizer et al. 2013).

Certain affordances, such as fitting through apertures, or into clothes, are determined by the size of one’s body. However, the processing of these affordances is determined not by body size directly but body size as *represented* by the body schema (Gadsby & Williams 2018). Given that, in AN, the body schema is oversized, it’s suggested that patients falsely judge such affordances: they judge that they are *too large* to accomplish certain actions, when in fact they are sufficiently small. Again, this is supported by evidence from experiments whereby patients are asked to judge whether they could accomplish certain size determined actions (Engel & Keizer 2017; Guardia et al. 2012; Guardia et al. 2010; Metral et al. 2014); misjudgement of these affordances suggests a similar phenomenon could occur throughout regular life. These misjudgements convey content such as “I couldn’t fit into her top” or “I can’t fit in between those chairs” (Gadsby 2017a, 12), providing patients with *false* information about their own body size (i.e. that it is *too large* for particular actions). These
affordance-based oversized experiences are proposed to further ground false beliefs about not yet meeting one’s ideal size.

A further step within empiricist models is to identify the neuropsychological abnormalities which give rise to the relevant experiences. In this regard, there is still much work to be done. False self-other and affordance judgments are suggested to arise from body image and body schema distortion, respectively—though it’s not yet clear the cause of such distortion (Gadsby 2017c). In terms of tracing spontaneous mental imagery back to a single cognitive or neurological factor, this task has yet to begin. There is still much work to be done before the empiricist model of false body size belief in AN is complete; nevertheless, I put this aside for now, in order to discuss a potential challenge to the account.

2.2. Searching for a second factor

While it seems likely that AN patients suffer from oversized experiences—providing evidence towards the belief “I am not my ideal size”—there is more to the story, in that such patients are also exposed to contradictory evidence, suggesting that their bodies are in fact thinner than their ideal size. This evidence comes in two varieties: patients have experiential evidence of their own body size through weight scale and clothes size readings; further, they are exposed to testimonial evidence from family, friends and clinicians attempting to convince them of their true body size.5

This evidential situation poses a particular kind of conundrum for the basic empiricist account: why, in the face of this disconfirmatory evidence, do patients maintain their delusional beliefs? Exposure to veridical evidence of body size should serve to dislodge the beliefs, at very least putting patients into a state of significant uncertainty regarding their own body size, yet in many cases this evidence appears to have little effect. Attempting to solve
this maintenance problem is ground well-trodden within empiricist research, whereby the most common strategy is to propose an additional cognitive deficit that delusional patients suffer from, beyond those that gives rise to the relevant abnormal experiences; these are generally termed “two-factor” accounts (Coltheart 2007; Davies et al. 2001).

Bolstering the demand for an additional cognitive factor is the findings of non-delusional analogues. These individuals suffer from the same abnormal experiences as delusional patients albeit lacking the associated beliefs (Coltheart 2007). For example, there are cases of patients with ventromedial frontal lesions who show the same lack of autonomic response as Capgras patients when presented with familiar faces (Tranel et al. 1995). Given this evidence, it’s assumed these patients have the same abnormal experience of familiar faces that Capgras patients do (recognition without affective response), despite lacking the associated beliefs (Davies et al. 2001, 144). Such cases are taken as further evidence that a second cognitive factor, beyond abnormal experience, is required. This factor would be present in delusional patients but absent in their non-delusional counterparts, dissociating the two groups.

There is some tentative evidence in favour of the existence of non-delusional individuals who undergo oversized experiences. For example, many healthy controls (especially those with high body concern) exhibit oversized perceptual body images (Baker et al. 1995; Plies & Florin 1992; Taylor & Cooper 1992). Similarly, in one study, experimenters found that roughly 50% of their healthy controls reported spontaneous mental imagery of an overweight body (Somerville et al. 2007, 439). These individuals might constitute non-delusional analogues, suffering from oversized experiences without the associated false beliefs. Although it’s too early for any definitive claims regarding the existence of non-delusional analogues, if indeed they do exist then this would constitute further evidence in favour of a second factor in the case of AN.
What could constitute this second factor, accounting for why AN patients maintain their beliefs in the face of contrary evidence and also, perhaps, dissociating them from non-delusional analogues? Earlier formulations of the two-factor account dictated a few essential properties of the second factor. First, it was suggested to be a domain general belief system dysfunction: a failure to inhibit pre-potent doxastic response, causing patients to unreflectively endorse their experiences as belief (Davies et al. 2001, 153). Second, its neurological basis is right frontal hemisphere damage (Coltheart 2007).

This early proposal is irreconcilable with the case of AN. Most obviously, there is no evidence to suggest AN patients exhibit right frontal hemisphere damage. Further, the domain general condition entails an unlikely consequence: that delusional patients should develop delusions in response to any strange experience e.g. perceptual illusions; yet there is no evidence that this is the case in AN—or any other delusion, for that matter (Davies et al. 2001, 153; Hohwy & Rajan 2012, 8; cf. Coltheart 2007, 1056).

Nevertheless, two-factor adherents often advocate a more general two-factor approach, which simply specifies two desiderata for any potential theory of delusion: answering how the content of the delusion arises and why the delusion is maintained (Langdon et al. 2008). Within this less stringent framework, the suggestion was made that, at least for some delusions, motivational biases might perform the second factor role (Davies 2009; Langdon et al. 2008; McKay et al. 2005; 2007). Under this proposal, the existence of motivational biases explain why the belief is adopted and maintained in the face of contrary evidence. Such a suggestion not only unshackles us from restrictive claims regarding neurological damage, it also avoids the aforementioned criticism against the domain general condition, as motivational biases only apply to certain kinds of incoming sensory input. Further, as I will demonstrate, this suggestion seems particularly compatible with the case of AN.
3. **Self-deception and biased evidence treatment**

Insight into the role that motivational biases might play as a second factor comes from the literature on *self-deception*—largely considered a paradigmatic example of individuals holding false beliefs as a result of motivational influences (Dewese-Boyd 2017). Self-deception occurs in two varieties: *straight* and *twisted*. In straight self-deception, one deceives oneself into believing something they desire to be the case, such as the parent who deceives themselves into believing their child isn’t experimenting with illicit drugs. This clearly doesn’t fit the case of AN—patients don’t desire to *not* be their ideal size, quite the opposite! Instead, AN has been suggested as an instance of twisted self-deception, whereby one deceives oneself into believing something one desires to be *false* (Sullivan-Bissett et al. 2016). As a non-pathological example, consider a jealous husband who, eagerly desiring his wife to be faithful, becomes convinced that she isn’t, despite all evidence to the contrary (Mele 2001).

In recent years, the relationship between self-deception and delusions has gained considerable attention (Bayne & Fernández 2009; Bortolotti & Mameli 2012; McKay et al. 2005). Of particular relevance is Mele’s (2001) *minimal* notion of self-deception. Mele claims that motivations can contribute to false belief through biasing the way in which individuals gather, attend to and interpret evidence (26-27). Such beliefs are, for Mele, instances of self-deception. That said, I won’t concern myself here with the conceptual issue of whether AN patients truly count as self-deceived. Such a proposal rests on much deeper debates regarding the philosophical underpinnings of the notion. Instead—following similar attempts aimed at other delusional beliefs (Davies 2009; Mele 2006)—I merely commandeer explanatory power from Mele’s analysis to reveal how motivational bias could constitute the second factor in the case of AN.
First, I review evidence of AN patients’ biases in evidence treatment. These biases, I suggest, cause evidence from oversized experiences to be amplified, while veridical body experience is neglected. Then, to proffer an explanation for how these biases arise, I adopt Mele’s (2001) proposal regarding the relationship between desires and hypothesis testing in self-deception.

3.1. Evidence Treatment Biases

3.1.1. Selective Evidence Gathering

Biased evidence gathering is considered to be a fundamental feature of AN. To start, the standard cognitive behavioural model of AN posits a repeated process of body checking (Fairburn et al. 1998). Aspen and colleagues (2013, 821) describe this process and its epistemic consequences:

Frequent checking of body parts (e.g., checking weight and/or the way clothes fit, intense scrutiny of particular [disliked] body parts; pinching skin to assess fatness) leads to strengthened AN-related behaviours. This repeated scrutiny of body parts serves as a confirmation bias in which individuals with AN seek out supporting evidence of their AN-related beliefs (e.g., ‘‘I am enormous’’)... This body checking constitutes a form of biased evidence gathering as it is targeted to reinforce false body size beliefs. Indeed, some patients even admit to engaging in such behaviour in order to “induce distress and hence increase their motivation to maintain dietary restraint” (Shafran et al. 2004, 100).
The standard cognitive behavioural model also claims that AN patients sometimes adopt *body avoidance* strategies:

Over time, the hypervigilant monitoring of shape and weight ('body checking') may become highly aversive. Some individuals are no longer able to tolerate the repeated self-examination. As a consequence they come to actively avoid monitoring altogether. *This avoidance maintains their shape and weight concerns in part because they no longer have a potential means of disconfirmation.* (Fairburn et al. 1998, 7, my emphasis)

Both these strategies, despite appearing contradictory, subserve the same evidential purpose (Tuschen-Caffier et al. 2015, 12). Body-checking behaviour is targeted at evidence which supports false body size beliefs (e.g. visually attending to “fatter” parts of the body, see section 3.1.2), while avoidance is targeted at evidence which disconfirms these beliefs.

Another form of evidence gathering bias relates to the *self-comparison* behaviour of ED patients. As discussed, ED patients are known to engage in self-comparison significantly more than healthy controls. In addition to this increase in frequency, there is a further bias to this self-comparison behaviour. In terms of self-comparisons, there are two possible kinds: upwards comparison, with those one believes are *better* than themselves (in this case, those with thinner bodies) or downwards comparison, with those one believes are *worse*. Not only do ED patients engage in significantly more self-comparison than neurotypical individuals, they engage in significantly more *upwards* self-comparison (Blechert et al. 2009; Corning et al. 2006; Green et al. 2009; Thompson et al. 1999). In other words, they actively seek out comparisons that will result in “they are thinner than me” judgments, reinforcing “I am not my ideal size” beliefs.
3.1.2 Attentional Biases

The belief confirming nature of patients’ evidence gathering practises is likely buttressed by biases at the attentional level. While much of the research into attentional biases in ED pertains to the study of semantic processing—e.g. through the use of Stroop colour-naming and dichotic listening tasks (for review, see Williamson 2000, 562-563)—most relevant to our current purposes is research from eye-gaze studies. This shows that when viewing themselves, those with eating disorders (or high eating disorder symptomatology) allocate their attention more towards self-identified “ugly” body parts (Freeman et al. 1991; Jansen et al. 2005; Tuschen-Caffier et al. 2015). Such biases likely work in tandem with body checking behaviour, specifically allocating attention to evidence which reinforces false body size beliefs.

In a similar vein, attentional biases are posited as relevant to the role of affordance based oversized experiences. Specifically, it’s claimed that an individual’s preoccupations contribute to the salience of certain affordances (Gadsby 2017a, 608). Consider that for any given environment-agent combination there are an infinite number of affordances that can potentially be processed. As such, the individual must filter these affordances, only attending to those which are personally relevant. Short-term desires (e.g. to engage in certain actions), of course, help to filter which affordances will be attended to but more general themes of preoccupation also play a role.

Consider the example of someone who obsessively thinks about skateboarding. The skateboarding related affordances of an environment (jumping those stairs, sliding down that rail) will have increased salience for such an individual: these affordances will jump out, commandeering greater attention. This is regardless of whether they indeed desire to
skateboard at that particular time. Similarly, for someone with less preoccupation with this general theme—thinking about it less often—salience of these affordances will be decreased, leaving them less regularly attended to and perceived. In this way, mental preoccupation in part determines affordance salience.

In the case of AN, patients exhibit intense preoccupation with ideas about their own body size (Mountjoy et al. 2014). As such, it has been suggested that this preoccupation leads to a heightened salience of body size related affordances (Gadsby 2017a, 609). In this way, size-related affordances which would otherwise remain unnoticed come to play a stronger role in grounding the relevant body size beliefs.

3.1.3 Interpretational Biases

There is also evidence of bias in how evidence is interpreted by AN patients. A common clinical observation is that they tend to infer negative meaning when presented with even slightly ambiguous information about their own body size. For example, “if someone comments to AN patients that they are ‘looking healthy,’ the patient may interpret that statement as ‘I am getting fat’” (Jackman et al. 1995, 342); similarly, feelings of fullness are often interpreted as “feeling fat” (Espeset et al. 2012, 523; Williamson et al. 2004, 714).

Biases in interpretation of ambiguous evidence have also been shown within experimental settings, through tasks which require participants to imagine themselves in situations described by ambiguous body size related sentences. Upon recall of the imagined situations, ED patients and those high in body dysphoria show a bias towards negative (i.e. “fat”) interpretations (Cooper 1997; Jackman et al. 1995; Williamson et al. 2000). Beyond the negative interpretation of ambiguous evidence, AN patients have also been known to explain away contradictory evidence. For example, Espeset and colleagues (2011, 183)
discuss a number of AN patients who interpret their low weight scale readings as evidence that their scales are broken, or that their bones are “lighter than usual”, rather than that their beliefs about not yet meeting their ideal body size are misguided.\(^8\)

Finally, interpretational bias is posited to play a role in the aforementioned empiricist model. Indeed, for spontaneous mental imagery to play a role in grounding body size beliefs, an additional interpretational bias must be present: the mental imagery must be interpreted as veridical. That said, it’s not yet clear how this bias might be characterised or even if it is present. It does, however, represent a potentially fruitful target for future empirical investigation.\(^9\)

### 3.1.4 Evidence treatment biases as second factor

In the preceding sections, I discussed a number of biases in the way AN patients gather, attend to and interpret body size related evidence (see figure 1). Evidence of these biases is nothing new and many have incorporated them into maintenance models of eating disorders before (Fairburn et al. 1998; Williamson et al. 2004). However, such biases are uniquely relevant to the proposed empiricist model: they solve the maintenance problem, accounting for why patients maintain their false body size beliefs despite being exposed to significant disconfirmatory evidence.

Specifically, under the proposed model, evidence treatment biases serve to emphasize evidence from oversized experiences and deemphasize evidence from veridical body size experiences. AN patients would actively seek out evidence from oversized experiences, while avoiding veridical body size experiences—a process which is further driven by attentional biases, whereby oversized experiences exhibit increased salience. Furthermore, interpretational biases would cause the evidence provided by oversized experiences to be
endorsed, while veridical body size experiences are disregarded or explained away. Such biases, in conjunction with the oversized experiences enumerated in section 2, present a persuasive two-factor account of how AN patients adopt and maintain their false body size beliefs.

3.2 The role of desire

3.2.1 The FTL model of hypothesis testing

While there is significant evidence of biases in evidence treatment in AN, a remaining question pertains to the cause of these various biases. Following other two-factor theorists, I proposed that motivational biases constitute the second factor. However, it’s not yet clear where motivations fit into the picture. If we take a fairly standard definition of motivational bias as: “[the influence on judgment by] the desirability or undesirability of events, consequences, outcomes, or choices” (Montibeller & Winterfeldt 2015, 1235), the task at hand is to show how desirability relates. This is where Mele’s (2001) theory of self-deception comes into play, the central insight of which is that biases in evidence treatment can arise from the undesirability of holding certain false beliefs.

Mele’s proposal sits within the context of what he terms the “Friedrich-Trope-Liberman” (FTL) model of hypothesis testing (Friedrich 1993; Trope & Liberman 1996). While we generally all desire to hold only true beliefs, the central insight of the FTL model is that holding some false beliefs is represented as less desirable than holding others. On this view, agents aren’t neutral truth trackers, steadfastly seeking true belief while avoiding falsity; rather, when it comes to belief formation, they seek to minimise “costly errors” (Mele 2001, 31).
In this model, errors are the false beliefs themselves and the cost of a false belief is “the cost, including missed opportunities for gains, that it would be reasonable for the person to expect the belief—if false—to have, given his desires and beliefs” (Mele 2001, 58). For example, falsely believing my child is experimenting with illicit drugs has a great cost associated with it, in terms of unnecessary psychological discomfort. Falsely believing the girl next door isn’t interested in me has a great cost also, in terms of missing out on a date. As such, falsely believing that my child is experimenting with drugs and falsely believing that the girl next door isn’t interested are both costly errors.

Now that the essential ingredients are laid on the counter, the best way to sample the flavour of this model is by contemplating how it might work in some typical instances of self-deception. Consider the examples of the pharmaceutically curious child and the romantically interested girl next door. As stated, falsely believing these statements come with costs. However, falsely believing the opposite of these—that my child isn’t experimenting with drugs and that the girl is interested—come with costs also: missing the opportunity to counsel the child and landing oneself in an embarrassing situation with the girl.

What determines which costs are greater? This comes down to the psychological profile of the individual. For people with certain psychological profiles—such as a significant adversity to psychological discomfort or an extreme desire to find a mate, coupled with extreme attraction to the girl next door—the costs associated with the former errors may outweigh the costs associated with the latter. When cost disparity occurs, hypothesis testing (i.e. evidence treatment) is influenced: one tests hypotheses in ways that minimise costly errors (Mele 2001, 41). This biased hypothesis testing in turn contributes to the adoption and maintenance of false beliefs.

**3.2.2 Application to the case of anorexia nervosa**
While the FTL model appears adequate in accounting for standard cases of self-deception—whereby biased evidence treatment results in individuals believing something they desire to be true—to be applicable to AN, it must also account for twisted cases, whereby the belief is something the subjects desire to be false. Fortunately, Mele (2001, chapter 5) offers a description for how the FTL model can also apply to the twisted variation.

In many putative cases of twisted self-deception, holding the relevant false beliefs—e.g. “my wife is unfaithful”, “I am not my ideal size”—undoubtedly causes psychological distress. As such, we might suppose that such beliefs would be classified as costly, causing the belief to be avoided. Yet in cases of twisted self-deception, Mele argues, cost disparity still occurs; for example, the costs associated with falsely believing “my wife is unfaithful” can be outweighed by the costs associated with falsely “my wife is faithful”. Mele describes how particular psychological profiles might result in such an imbalance:

It certainly is conceivable that, given a certain psychological profile, a strong desire to maintain one’s relationship with one’s spouse plays a role in rendering the potential error of falsely believing one’s spouse to be innocent of infidelity a “costly” error, in the FTL sense, and more costly than the error of falsely believing one’s spouse to be guilty. After all, the former error may reduce the probability that one takes steps to protect the relationship against an intruder. (2006, 114)

A similar situation may hold in the case of AN. Undoubtedly, falsely believing that one isn’t their ideal size causes psychological discomfort but falsely believing one is their ideal size may be regarded as even more costly. Indeed, given the strong social element involved in AN patients’ drive for thinness (i.e. the desire to be seen as thin) a situation in which one wasn’t
thin but believed they were would be significantly undesirable and therefore classified as significantly costly. Such a situation would render one unwilling to amend their current state (dieting to decrease size) due to ignorance of their own body size—unbeknownst to them, all their peers would see them as overweight.

In addition to costs derived from the desire to not be seen as overweight, the pathological desire to not be overweight might also play a role, through lowering the costs associated with falsely believing that one hasn’t reached their ideal size. Indeed, falsely holding such a belief is in many ways advantageous for someone with a pathological desire for thinness: it helps to motivate weight loss efforts, ensuring that one never slips accidently into being overweight. This final point coheres with the aforementioned testimony from AN patients who claim they often engage in biased evidence gathering in order to increase motivation for dietary restriction (Shafran et al. 2004, 100).

Given these associated costs and the relative imbalance between them, the FTL model predicts that such agents will engage in biased forms of hypothesis testing, a natural result of the tendency to avoid costly errors. Further, we might suppose that the greater the cost imbalance (as might be the case in AN), the more biased the hypothesis testing practises become. Given the pathological nature of patients’ desire to neither be overweight nor be seen as overweight, we might thus expect significant evidence treatment biasing to occur. To conclude, the FTL model offers an explanation for how AN patients’ own fears about being overweight bias their evidence treatment practices, via a high cost associated with falsely believing “I am my ideal size” and a low cost associated with falsely believing the opposite.

It seems plausible that individuals who pathologically value thinness might consider falsely believing they are their ideal size to be vastly more costly than falsely believing the opposite. However, we needn’t rely on speculation alone here: the issue is also open to empirical investigation. Specifically, questionnaires could be designed which aim at
establishing the costs patients associate with holding certain false beliefs. While offering a specific proposal for what form such questionnaires would take is beyond the purview of this essay, one possibility is that participants are read short vignettes—whereby an individual holds the relevant false belief—and asked to rate how undesirable such scenarios would be. It might even be found that differences in undesirability ratings go some way towards accounting for the differences in belief conviction between patients (Phillipou et al 2017).

Of course, confirming that AN patients exhibited the relevant cost imbalances would not settle the issue: there is still the matter of empirically validating the FTL model itself and confirming that costs associated with false belief outcomes do indeed cause biased evidence treatment, of the kind AN patients exhibit. Nevertheless, the suggested experimental approach certainly represents a step in the right direction.

Within the context of the FTL model, I have suggested that AN patients’ strong desire to neither be overweight nor be seen as overweight may result in a high cost association with falsely accepting the hypothesis “I am my ideal size” and a low cost association with the opposite. According to this model of hypothesis testing, evidence treatment is biased towards avoiding costly false beliefs. This explains how the biases in evidence gathering, interpretation and attention reviewed in section 3.1 might arise.11 This (pathologically) strong desire to neither be overweight nor be seen as overweight might then be said to constitute the second-factor: causing biased evidence treatment practises, contributing to the maintenance of patients’ false body size beliefs. Similarly, this strong desire may dissociate AN patients from non-delusional analogues who undergo similar oversized experiences, without the associated body size beliefs. While this proposal is in need of appropriate empirical validation, it appears credible as an explanation for how patients’ treatment of body size related evidence comes to be biased.
4. Conclusion

While the empiricist approach demonstrates promise as an explanation for the false body size beliefs of AN patients, as with other delusions, a second factor is needed. Following a common proposal within two-factor research, I suggested that motivational biases might fill this role. Specifically, I proposed that biases in evidence treatment cause oversized experiences to be amplified, while veridical size experiences are discounted and, further, that these biases result from motivational influences—via cost imbalances related to beliefs about meeting one’s ideal body size.

It remains to be seen how this model can be applied to other eating disorders. The most likely candidate is, of course, BN. Not only do the two disorders frequently overlap (Vitousek et al. 1998, 396) but BN is also associated with perceptual body image distortion (Norris 1984; Whitehouse et al. 1986), recurrent spontaneous mental imagery (Somerville & Cooper 2007; Somerville et al. 2007) and systematic biases in body size evidence treatment (Williamson et al. 2004). Despite a relative lack of body representation and delusional belief research targeting BN, it seems likely that the same model will apply to many of these patients, even if a few modifications are needed. Ultimately, I suspect that the kinds of behavioural (and weight) differences that distinguish AN from BN patients won’t prove relevant to whether the proposed model applies; rather, the important factor will be the existence and prevalence of the relevant beliefs. As such—and in following with the cognitive neuropsychiatric tradition—it may prove beneficial to overlook the apparent diversity in the conditions themselves and instead narrow our focus onto the beliefs.

Even restricting our scope to AN, there is still much work to be done. While some of the discussed biases in evidence treatment in AN are supported by robust bodies of evidence, others clearly need further empirical validation and some are little more than tentative
suppositions. Further, both the FTL model of hypothesis testing and the costs patients associate with holding certain false beliefs call for meticulous empirical probing. Finally, there might be other factors at play here—perhaps some yet to be identified cognitive deficit(s)—which contribute to the maintenance of these false beliefs. It’s still early days in terms of this empiricist approach, with disagreement and debate undoubtedly on the horizon. Nevertheless, once the dust settles and a finalised model is left standing, it seems likely that motivational biases will play a decisive role.

1 Herein I adopt the mainstream assumption that such delusions are indeed beliefs (Bortolotti 2010; Clutton 2018; cf. Currie 2000; Gerrans 2014)
2 Evidence of this comes from figural drawing scale experiments: participants are asked to select body size silhouettes that match both their current and ideal body sizes, with the difference in size taken to be an indication of “body dissatisfaction” (Moussally et al. 2017).
3 This isn’t to emphatically state that the oversized experiences discussed are in no way related to the deficits in interoceptive processing that have been uncovered. A link between these two forms of dysfunction may emerge, though it’s not yet clear what this link might be.
4 Unlike with the other forms of oversized experiences, most of the research into spontaneous mental imagery in ED has been conducted on BN, rather than AN, patients (cf. Cooper et al. 1998). This lends further support to the possibility that the model discussed here might apply to a number of eating disorders (see section 4).
5 It might be that patients also experience their body size accurately through direct visual perception, though this is a source of some contention. While some patients claim they (directly) see themselves as thin, others claim the opposite (Epes et al. 2011). Based on evidence from mirror exposure research, it has recently been argued that AN patients’ direct perception of their bodies must be veridical (Gadsby 2017c, 27). Nevertheless, this is still an open question (cf. Mohr et al. 2016)
6 Within delusion literature, the maintenance problem just posed is sometimes distinguished from the adoption problem, which requires an explanation for why the abnormal content was adopted as belief in the first place (Davies & Egan 2013). That said, I won’t delve into the specifics of this distinction here as, although interesting, it’s orthogonal to my central thesis (cf. Gadsby 2017b, 501-503).
7 In more recent two-factor literature, there is some debate over whether indeed there is a conscious experiential element to this autonomic dysfunction; I won’t delve into those debates here though (see: Coltheart et al. 2010; Young 2011).
8 This example coheres with research into the confabulatory practices of delusional patients, who often arrive at patently implausible explanations for evidence which conflicts with their delusional beliefs (Langdon & Bayne 2010, 323).
9 See Holmes and Mathews (2010, 354-355) for a discussion of some different hypotheses for why clinical patients might come to interpret spontaneous mental imagery as veridical.
10 An interesting point arises here regarding whether the relationship between desires and biased hypothesis testing must be consciously mediated by a belief that this form of hypothesis testing will avoid the relevant costly error (Mele, 2001, 31-32, 42-46). In some cases, such as the mentioned excerpt, patients clearly are aware that certain evidence treatment practices (i.e. body checking) will aid in avoiding undesirable situations. Yet this needn’t be the case with all instances of biased hypothesis testing. For example, it seems less likely that attentional and interpretational biases are consciously mediated and indeed the FTL model allows that much of this biasing is “automatic and inflexible … reflecting the operation of evolved cognitive adaptations to a range of biologically significant problems” (Friedrich, 1993, p. 317).
It’s worth highlighting that this story would markedly differ from the discussed hypothesis regarding affordance salience in AN. While that hypothesis claims all size-determined affordances have increased salience (due to patients’ mental preoccupation with body size related themes), the proposed self-deception hypothesis suggests a particular subset of size-determined affordances (those likely to reinforce beliefs about being overweight) would exhibit increased salience. Such a bias would manifest in patients attending to affordances they believed their bodies were too large for.
References


doi:10.1080/13546800444000074


Metral, M., Guardia, D., Bauwens, I., Guerra, M., Lafargue, G., Cottencin, O., & Luyat, M. (2014). Painfully thin but locked inside a fatter body: abnormalities in both
anticipation and execution of action in anorexia nervosa. *BMC research notes*, 7(1), 707.


doi:10.1162/jocn.1995.7.4.425


Figures

<table>
<thead>
<tr>
<th>Evidence Gathering</th>
<th>Attentional</th>
<th>Interpretational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active body checking/avoidance</td>
<td>Visual biases towards disliked body parts</td>
<td>Negative interpretation of ambiguous evidence</td>
</tr>
<tr>
<td>Upwards self-comparison</td>
<td>Heightened affordance salience</td>
<td>Rationalising contradictory evidence</td>
</tr>
</tbody>
</table>

Figure 1.