

Forthcoming in *Philosophical Psychology* – Please cite published version

What Makes Delusions Pathological?

Valentina Petrolini¹

Bortolotti (2009) argues that we cannot distinguish delusions from other irrational beliefs in virtue of their epistemic features alone. Although her arguments are convincing, her analysis leaves an important question unanswered: What makes delusions pathological? In this paper I set out to answer this question by arguing that the pathological character of delusions arises from an executive dysfunction in a subject's ability to detect relevance in the environment. I further suggest that this dysfunction derives from an underlying emotional imbalance – one that leads delusional subjects to regard some contextual elements as deeply puzzling or highly significant.

Introduction

In *Delusions and Other Irrational Beliefs* (2009), Lisa Bortolotti puts forward two important claims concerning delusions and their relationship to norms of rationality. First, she gives reasons to pry apart the criteria for belief ascription and the criteria for rationality, arguing

¹ Valentina Petrolini is a PhD candidate in philosophy at the University of Cincinnati. Correspondence to: Valentina Petrolini, Department of Philosophy, University of Cincinnati, 2700 Campus Way, Cincinnati, OH 45221, USA. Email: petrolva@mail.uc.edu

* Special thanks to Peter Langland-Hassan for the helpful, patient and always encouraging feedback on multiple versions of this paper. I am grateful for the comments I received from Heidi Maibom, Alexander Jeuk, Alberto Arati and Carlotta Facini on different aspects of this work. Thanks to the audiences at the *Society for Applied Philosophy Conference* (Sara Protasi and Justin Caouette in particular), the *Collective Intentionality IX* conference and the *European Society for Philosophy and Psychology* conference, where I presented some of this material. I would also like to thank three anonymous referees for their detailed and engaging comments on previous versions of this paper. This work was supported in part by a summer fellowship from the TAFT Research Center.

Forthcoming in *Philosophical Psychology* – Please cite published version

that the process of ascribing a belief should not be constrained by whether the belief is rational (thereby rejecting the Rationality Constraint Thesis, or RCT). Second, she argues that delusions share their distinctive epistemic features – i.e. their failure to live up to rational norms – with other irrational beliefs. This moves her to characterize all beliefs as lying on a continuum of rationality, with delusions at the most irrational end of the spectrum (Continuity Thesis, or CT). Even if we accept her conclusions, Bortolotti’s analysis leaves an important problem unsolved: if delusions cannot be distinguished from other irrational beliefs by their failure to live up to rational norms, what makes delusions pathological? In this paper I set out to provide a solution to this problem – one which provides clear links between the irrationality of delusions and their status as pathological conditions deserving of clinical intervention.

The paper is divided into three sections: in the first I briefly reconstruct Bortolotti’s theses and I propose a way in which her account of delusions can be amended. In the second I suggest fine-tuning the notion of *rationality* beyond the following of procedural, epistemic and agential norms, in order to include the ability to detect *relevance* as a crucial cognitive capacity. This addition benefits Bortolotti’s account because it characterizes delusions as pathological not merely in virtue of their violation of canonical norms of rationality. Instead, I suggest that delusions arise from a subject’s difficulty in appropriately detecting relevance in the environment, and *this* leads to violations of rational norms. In this section I also draw on recent studies (Ibanez Casas *et al.*, 2013; Grover *et al.*, 2011) to show that relevance detection can be measured empirically through the more familiar notion of *executive functions*. In the third section I render this aberrant process of relevance detection more precise by focusing on the role played by *emotions* in the experience of delusional subjects. More specifically, I argue that delusional subjects – and schizophrenic patients affected by delusions in particular – present a significant degree of emotional dysfunction that causes *and* accompanies the executive disruption. By doing

Forthcoming in *Philosophical Psychology* – Please cite published version

so I characterize relevance detection as an affectively laden process, where emotions work as cues directing our attention to specific elements of the context. In this sense, delusional subjects can be described as individuals who exhibit an exaggerated emotional response to specific environmental stimuli and – as a consequence – perceive them as overly threatening or puzzling.² Therefore, whereas Bortolotti’s proposal leaves us without an account of what makes delusions distinctively pathological, I argue that they can be successfully characterized as *executive* and *emotional dysfunctions*.

1. Bortolotti’s Proposal

In *Delusions and Other Irrational Beliefs*, a great part of Bortolotti’s analysis is devoted to the arguments against RCT. Indeed, her main preoccupation lies in determining whether clinical delusions should be regarded as “genuine instances of beliefs”, despite their patent irrationality (Bortolotti 2012b, p. 1). The argument is roughly the following: we cannot deny belief status to delusions by appealing to the fact that they violate norms of rationality, because many ordinary beliefs can be charged with the same kind of violation. Indeed, Bortolotti shows that delusions and ordinary beliefs share crucial epistemic features and often fail to fully meet the standards of three canonical norms of rationality: procedural, epistemic and agential.

The norms of *procedural* rationality require a certain degree of coherence or good integration among beliefs, and most delusions indubitably lack this feature: for example, subjects affected by persecutory delusions often fall prey to loose associations and tend to see disparate events as closely connected with one another. However, this lack of harmony and integration

² Although the nature and extent of emotional disturbances in psychotic patients is still poorly understood – see Antievic *et al.* 2012 – some recent studies suggest that they would ascribe affective importance to stimuli that controls regard as neutral (Hall *et al.* 2008) as well as interpret facial expressions as more threatening than they are (Surguladze *et al.* 2006).

Forthcoming in *Philosophical Psychology* – Please cite published version

among beliefs also occurs in many non-pathological cases. For example, superstitious beliefs are often badly integrated: a doctor working in emergency room can be committed to the principles of science and at the same time be convinced that a night of full moon causes more accidents (Bortolotti 2009, p. 85). The norms of *epistemic* rationality focus instead on the relationship between beliefs and evidence, requiring a certain degree of responsiveness in light of supporting or disconfirmatory events. Bortolotti acknowledges that, paradigmatically, delusions are characterized as beliefs not well-supported by evidence or extremely resistant to counterevidence. For example, patients affected by erotomania strenuously defend the belief that someone is in love with them by marshaling insufficient or completely unrelated evidence. However, a similar degree of resistance to contrary evidence can also be detected in more ordinary cases such as religious beliefs or racial prejudices: indeed, these are often sustained despite the lack of any empirical support and tend to be defended through confabulation or faulty reasoning (Ibid., pp. 121 & 150). Finally, the norms of *agential* rationality concern the connection between beliefs and actions, and require that the subject shows some behavioral endorsement of a proposition in order for a belief to be ascribed to her. In fact, delusional individuals sometimes display behaviors that appear significantly different from the ones that would be expected if they did believe the content of their delusions. A striking example is the Capgras syndrome, in which some patients claim that their spouse has been replaced by an impostor without taking any further action (e.g. calling the police, looking for the missing person). Yet, Bortolotti also points out a series of ordinary life situations in which the agential rationality constraint seems violated. Good examples are cases of behavior-attitude inconsistency, such as acknowledging the importance of condoms for HIV prevention while systematically avoiding using them in practice (Ibid., pp. 172-173).

All these examples show that deviations from the norms of procedural, epistemic and agential rationality “are not the exception to the rule, but widespread and systematic” (Ibid., p.

Forthcoming in *Philosophical Psychology* – Please cite published version

78). Thus – Bortolotti argues – we should abandon the rationality constraint on belief ascription because both delusions and ordinary beliefs blatantly fail to satisfy these norms. A consequence of this analysis seems to be that the difference between delusions and ordinary beliefs should not be regarded as a difference in kind, but rather as a difference in *degrees* of rationality (CT). For the line between abnormal and normal cognition does not look to be clear-cut: clinical delusions appear “typically irrational to a greater extent or irrational across more dimensions than non-delusional beliefs, but they are irrational roughly in the same way” (Bortolotti 2012b, p. 39).

Yet, this conclusion leaves us with a puzzle. Once it is granted that delusions cannot be distinguished from other beliefs (only) in virtue of their irrationality, we are still left with the problem of identifying what it is that is *pathological* about them. Bortolotti seems willing to grant this point at the end of her book: “What makes delusions pathological? Whatever it is, it is not their being irrational, because the irrationality of delusions is not different in kind from the irrationality of everyday beliefs” (2009, p. 259). Then she offers a few remarks on what could make delusions pathological: on the one hand, they undoubtedly affect the health and well-being of the subjects who report them, often suffering from high levels of distress and social isolation; on the other, they may be more puzzling than other irrational beliefs, to the extent that we often have a hard time understanding them (Bortolotti 2012a, pp. 52-53). However, neither of these suggestions looks satisfactory. Indeed, there are forms of superstition that may affect a person’s well-being to a significant degree, for example by preventing her from leaving the house on some “unlucky” day of the week. Again, the endorsement of some socially unaccepted beliefs – such as extreme forms of racism or sexism – may cause various degrees of social isolation. Moreover, it is not always true that delusions present a more puzzling content than other irrational beliefs: thinking that my partner is cheating on me (Othello syndrome) seems far more plausible than believing that a full moon is responsible for more car accidents.

Forthcoming in *Philosophical Psychology* – Please cite published version

Therefore, one way to amend Bortolotti's account would be to explain the pathological nature of delusions in a way that does not depend on the *mere* violation of rationality constraints. Indeed, even granting that delusions lie on a continuum with ordinary beliefs leaves her proposal underspecified and relying on the vague consideration that delusions violate rationality constraints "to a greater extent" or "across more dimensions" than ordinary beliefs. As Bortolotti puts it: "The degree of irrationality tracks *both* how much [delusions] deviate from norms of rationality for beliefs *and* how many norms of rationality they deviate from" (Bortolotti 2012b, p. 4). In this passage the tension within Bortolotti's account becomes more apparent. On one hand, she admits that epistemic features – i.e. adherence to rational norms – cannot successfully demarcate delusions from other irrational beliefs. On the other, she draws this distinction by appealing to the extent to which delusional beliefs deviate from canonical norms or even to the number of norms that they happen to violate. On this view it becomes difficult to characterize delusions in terms of their distinguishing features, because the appeal to norms of rationality does not discriminate among relevant cases. Indeed, some ordinary beliefs might be irrational to the greatest possible degree or across more dimensions without thereby qualifying as delusional. For example, the religious belief in resurrection seems at the same time practically impossible (procedural), not supported by evidence (epistemic) and problematic to endorse in practice (agential).

More recently, Bortolotti (2015a & 2015b) has stressed the idea that delusions have significant epistemic benefits and thus may qualify as "adaptive misbeliefs". Indeed, the endorsement of a delusional belief often performs a defensive function and prevents a subject from accepting other beliefs that would cause more serious harm or distress. For example, she discusses the case of a patient who develops reverse Othello syndrome – i.e. the belief that your partner is faithful when s/he is not – after having been abandoned by his girlfriend following a car

accident that had left him permanently disabled. In such a situation, the endorsement of the delusional belief – e.g. “My partner still loves me” – can be described as an adaptive response that helps the patient cope with the situation without falling prey to depression and anxiety (2015a, p. 4). At the same time, Bortolotti acknowledges that delusions carry significant costs that may outweigh benefits (Ibid., p. 7). However, she argues, we may want to see delusions as *epistemically innocent* as opposed to merely adaptive or maladaptive. In order to qualify as epistemically innocent, a belief needs to satisfy two conditions: a) The subject acquires some significant epistemic benefits by endorsing it; b) other beliefs that would confer similar benefits are not available to the subject at the time. Now, even granting a) in virtue of the benefits that Bortolotti highlights – e.g. anxiety reduction – this formulation puts a lot of pressure on b) and renders the question about pathology even more significant. In other words: What *makes it the case* that some subjects regard the delusional explanation as the best (or the only) option available? Bortolotti’s answer here is tentative: whereas she mentions several limitations such as “reasoning biases, need for closure and motivational charges”, she also acknowledges that none of these factors constitute a marker for mental illness since they are all widely distributed in the non-clinical population (2015b, p. 16). Again, while the emphasis on continuity reduces the distance between pathological and ordinary cognition, it renders the question of what makes delusions clinically interesting more pressing. Indeed, if delusions are to count as “innocent”, the way in which alternative, non-delusional beliefs are unavailable to a subject needs to be strong. In other words, a convincing etiological pathway has to be provided to clarify why some delusional explanations come to be regarded as the best – or only – options available.³

In sum, Bortolotti’s arguments leave us with a problem. First she shows that delusions are similar to ordinary beliefs in terms of their relation to norms of rationality, but fails to identify

³ I would like to thank two anonymous referees for allowing me to make this point more explicit.

Forthcoming in *Philosophical Psychology* – Please cite published version

what makes delusions pathological (2009 & 2012). Indeed, pointing to a general decrease in well-being is not sufficient, because we do not know what about delusions may be specifically responsible for such a decrease. Then she argues that delusions confer significant epistemic benefits in situations where our choices are constrained, but does not explain why some people fall prey to delusions while others do not (2015a & 2015b). If this is correct, we need some other way of capturing what it is that makes delusions pathological and that also explains why delusional subjects have the right to be medically treated. After all, individuals affected by delusions often exhibit dramatic levels of impairment in their well-being and social functioning, as well as a vast array of affective disturbances such as depression, anxiety and negative symptoms (e.g. anhedonia, flat affect). One reaction to Bortolotti's argument would be to conclude that the pathological features of delusions have little or no relation to their epistemic features, and to view those features as a distraction in the search for the cognitive basis of delusions. However, I want to pursue a different path – one which accepts Bortolotti's points while still drawing meaningful connections between the epistemic features of delusions and their status as pathologies. This will amount to giving an account of the specific cognitive abnormalities specific to delusions that, in the vast majority of cases, lead to the impaired well-being characteristic of pathology.

In the second and third sections of this paper I set out to accomplish this goal. First, I show that the notion of rationality can be fine-tuned in important ways beyond the three canonical norms. In particular, I introduce the notion of *relevance detection* and argue that delusional subjects exhibit a specific dysfunction that affects their ability to properly pry apart relevant and irrelevant stimuli in the environment. In this sense I propose to use disturbances in relevance detection – as opposed to violations of norms of rationality – as one of the proper marks for delusions. Second, I explore the influence of *emotions* on the reasoning of delusional

Forthcoming in *Philosophical Psychology* – Please cite published version

subjects, suggesting that the dysfunctions in relevance detection could derive from underlying emotional imbalances that give rise to a peculiar experience of reality. In a nutshell, I argue that the pathological nature of delusions can be characterized in terms of executive and emotional dysfunctions.

2. Beyond the Canonical Norms: Relevance Detection

2.1. What is Relevance Detection?

The norms of procedural, epistemic and agential rationality discussed by Bortolotti already encompass many features that we normally attribute to a subject when labeling her as “rational”: coherence or good integration among beliefs, responsiveness to evidence, and some connection between beliefs and action. Here I suggest fine-tuning the notion of rationality in order to include another capacity – *relevance detection* – that plays an important role in our cognitive makeup while often preceding the application of the norms discussed by Bortolotti. In this sense, relevance detection captures the same phenomena described by the norms at a different level of description. Relevance detection can be broadly characterized as the ability to “tag” specific environmental elements as significant according to an individual’s interests and goals. In other words, it works as a filter that allows an individual to focus her attention on certain stimuli while disregarding others. What makes a piece of information more or less relevant depends on the context as well as on the individual’s current goals. For example, if a friend is venting with me about a pressing matter, I would probably attend to his voice and facial expression more than to what is hanging on the wall behind us. Yet, if my friend is an artist who is outraged about the fact that his masterpiece has not been included in the local exhibition, attending to the painting on the wall becomes of the utmost importance.

Notably, this capacity often operates automatically, in a non-deliberative and non-inferential way. For this reason, it makes sense to think that relevance detection could have evolved as an ability to facilitate cognitive systems to achieve their basic goals in an expedient and time-efficient way. As some researchers have suggested, this would explain the high degree of salience exhibited by certain stimuli – such as angry or fearful faces – that have been shown to elicit immediate responses even in cross-cultural settings (see Sander, Grafman & Zalla, 2003). Relevance detection can thus be described as the ability that allows a system to determine which environmental stimuli are important for the achievement of its current goals. A disturbance in this ability could thus lead to impairments in cognitive processes such as data collection and interpretation, which appear particularly skewed in delusional cases (see Speechley & Whitman 2010).

With respect to our previous discussion, relevance detection relates to each of the norms discussed by Bortolotti by capturing similar phenomena at a lower level of description, and may thus prove more helpful than the norms in determining what goes wrong in the experience and reasoning of delusional subjects. In terms of *procedural* rationality, what we find relevant in assessing whether we should accept a belief may very well depend on the level of the belief's integration with other beliefs. This sort of top-down influence of higher level cognitive states – such as beliefs – on perception has been widely supported by empirical data. For example, the commonsense belief that larger objects weigh more than smaller ones has been shown to influence our perception of an object's mass (see Lupyan 2015). Yet, there are many situations in which this relationship works in the opposite direction. That is, an alteration of relevance detection may be responsible for the revision of our beliefs, rather than the other way around. For example, if I live in a safe neighborhood and I believe that it is very unlikely for me to get robbed while walking home at night, I am somehow entitled not to pay attention to details such as the

lack of light or the presence of other people on the street – e.g. I do not need to check which side of the street is more illuminated so that I can decide where to walk accordingly. However, if I hear about a friend who got robbed in a different neighborhood, my walking-home behavior may change even if my beliefs about the safety of the area are maintained. Suddenly, I may start to pay attention to details that I had previously overlooked – e.g. the store’s closing time – and my habits can be modified to the extent that an inconsistency between two beliefs finally ensues.⁴ As a result, the application of a procedural norm may take place and the previous belief – i.e. “My neighborhood is safe” – may be updated according to the new context in order to avoid inconsistencies – i.e. “My neighborhood is not safe anymore.” Here, it is important to note that the revision does not take place at the high level of rational norms, where the subject weighs one belief against another. Rather, integration among beliefs is the consequence of some alteration in relevance detection that unfolds over time and can have multiple causes: e.g. something I heard from a friend, or my increased level of anxiety. In this case, relevance detection is needed for procedural norms to be applied. This also helps us to see pathological cases in a new light: for example, patients affected by anosognosia who exhibit “bad integration” or commit to “conflicting reports that cannot both be true” (Bortolotti 2009, pp. 64-65), may do so *because* they experience some anomaly in relevance detection. That is, they may hold fast to some previous belief – e.g. “I am not paralyzed” – and yet relating to external stimuli in a way that goes against this conviction – e.g. avoid lifting bags or climbing stairs.⁵ In this sense, characterizing the issue only in terms of breakdown of procedural rationality does not address the problem at the appropriate level. Indeed, in many cases – both ordinary and pathological – the

⁴ This example is adapted from Elgin 2007.

⁵ Similarly, cognitive models – such as the one put forward by Freeman, Garety and their collaborators (2002; 2001) – see delusions as arising from a disturbance in attributing relevance. For example, patients affected by persecutory delusions tend to experience neutral or even positive events (e.g. a passing glance, a smile) as threatening (e.g. the glance is taken as evidence of plotting, the smile as a nasty one) – see Freeman *et al.* 2001, p. 336.

integration among beliefs importantly depends on a system's ability to detect relevance in the environment.

A similar relationship can be found between relevance detection and *epistemic norms*. Surely, the ability to detect relevance in a context is closely connected with the way in which an individual responds to an available body of evidence. Bortolotti herself acknowledges that bottom-up effects of this kind are frequent, as she mentions the way in which “selective attention” to specific stimuli contributes to reinforce a subject's delusional interpretation of “objectively neutral events” (Bortolotti 2009, p. 140). Yet, she still characterizes this process in a way that privileges a high-level of description: for instance, she circumscribes epistemic norms to the “gathering and weighting of evidence”, and talks about the way in which subjects “insulate certain beliefs from potentially disconfirming evidence” (Ibid., p. 116). Even her treatment of the notion of selective attention reflects the commitment to a top-down view: “Beliefs guide the way in which we gain information, so that the information we process is more likely to reinforce our beliefs” (Ibid., p. 153). Again, this conception works well in certain cases: for example, when sexist beliefs – e.g. “Women are bad drivers” – selectively guide the way in which we collect evidence – e.g. by paying more attention to the mistakes women make behind the wheel. Yet, in many other contexts beliefs seem to emerge from, rather than guide, the way in which various pieces of information are gathered and interpreted. Think about a person walking into a party and observing a group of people laughing: here, the formation of the belief “They are laughing at me” as opposed to “Someone just made a funny joke” surely depends on how some contextual cues are interpreted – e.g. how the people looked at the person, when they stopped laughing, how long did the silence last. In these cases, the ability to detect relevance importantly *precedes* the application of epistemic norms. Indeed, a process of filtering has to take place in order to determine which pieces of evidence support or disconfirm a belief: relevance detection acts as

such a filter by tagging some pieces of evidence as more important than others. This is (again) compatible with the clinical evidence marshaled by cognitive accounts of delusions, in particular with the idea that delusional reasoning would involve biases toward the collection of confirmatory evidence. In particular, delusional subjects would prioritize certain stimuli over others – e.g. potential threats – and would require fewer instances than controls to endorse a certain explanation – i.e. “jumping to conclusion” bias (see Garety *et al.* 2001). As I discuss in the next section, these processes of tagging and filtering are deeply influenced by emotions. In this sense, operations such as the ones described by Bortolotti – i.e. “weighing evidence” or “insulating beliefs from potentially disconfirming evidence” – can take place *because* our cognitive system has already regarded an event or object as relevant or worth our attention. Clearly, things are not this simple and feedback loops between top-down and bottom-up effects often occur. In the last example, detecting the sudden silence in the room as particularly relevant may in turn result from (consciously accessible) beliefs about myself and my social abilities. For instance, if I already believe I am unattractive, or not worth having a conversation with, I would surely be more inclined to interpret contextual cues in a certain way, making the evidence for a certain belief – e.g. “They are laughing at me” – much easier to collect.⁶ Therefore, the distinction between relevance detection and epistemic norms appears particularly subtle: the former captures the subject’s ability to become aware of what may count as evidence in a given context; the latter describes the various ways in which a subject can use a body of evidence at her disposal, once she has become – more or less – aware of it.

Finally, relevance detection and norms of *agential rationality* are also connected to one another in a similar way. Surely there are situations where the goal of preserving coherence between beliefs and actions influences the way in which things appear more or less salient to us.

⁶ I would like to thank an anonymous reviewer for expanding on this example and allowing me to clarify this point.

Forthcoming in *Philosophical Psychology* – Please cite published version

For example, if I believe I should lose weight and thus try to avoid eating sweets, the croissants I see at the coffee shop seem particularly relevant while in normal circumstances they would go unnoticed. Yet, in many other cases the relationship between a person's beliefs and her behavior is influenced by what appears relevant in the current context. For example, I may come to believe that a colleague needs help and then act upon that belief because I detect an anomalous behavior on her part: perhaps she has a strange smile and puffy eyes every morning, or she turns down coffee one time too many, or comes to the office hungover on a Tuesday. More generally: in some cases the connection between beliefs and behavior guides our detection of relevance in the environment, like when we are on a diet and every piece of cake we see looks appealing. Yet, in other cases relevance detection plays a role in determining how beliefs and behavior interact, thus importantly preceding the application of agential norms. In the case of the alcoholic colleague, I first detect some environmental cues, then form the belief that the person needs help, and finally act upon my belief and direct her to a professional counselor. Once again, the relationship between relevance detection and agential norms is complex and includes top-down and bottom-up effects.

Let us take stock: so far I have argued that Bortolotti's discussion of procedural, epistemic and agential norms can be significantly refined in order to include the ability to detect relevance in a context. In particular, I have shown that in some cases relevance detection importantly precedes the application of canonical norms, while in others it allows us to capture similar phenomena at a lower level of description. In this sense, the picture of belief formation put forward by Bortolotti – despite its correctness – tells us only one half of the story: while she focuses on how individuals follow or violate rational norms at a conscious, attentive level, she neglects an array of cases where lower-level, non-deliberative activities such as relevance

Forthcoming in *Philosophical Psychology* – Please cite published version

detection play a crucial role.⁷ In the next section I focus on relevance detection from this broadened perspective: first I show that this ability can be measured empirically through the notion of *executive functions* (§2.2). Then I suggest that, due to emotional irregularities, delusional subjects exhibit a particular form of executive impairment (§2.3).

2.2. How do we Measure Relevance Detection?

The crucial role played by relevance detection in our cognitive makeup has been explored in neuropsychiatry and clinical psychology through the analysis of the so-called *executive functions*. These abilities have been now separated from other cognitive capacities such as memory and language, and are strongly associated with specific brain areas: in particular, they seem to involve the prefrontal cortex and its basal ganglia-thalamic connections. Executive functions encompass a broad range of cognitive skills whose role seems twofold: on one hand, they are associated with high-level abilities – e.g. abstraction or planning – while on the other they are responsible for monitoring complex behavior. Most researchers divide executive functions into three sub-capacities: *shifting* is defined as the ability to switch between different tasks or different aspects of the same task; *update* as the ability to incorporate relevant information while removing non-relevant ones; *inhibition* as the capacity of filtering out inappropriate responses.

The function of *shifting* seems particularly important for our purposes because the ability to switch between different mental scenarios involves “the disengagement of an *irrelevant* task set and the subsequent active engagement of a *relevant* task set” (Miyake *et al.*, 2000, p. 55.

⁷ As I show in the party example above (pp. 12-13), this picture may be overly simplistic: indeed, it might be difficult to disentangle top-down and bottom-up effects in any given situation. Most cases would probably feature complex feedback structures, where pre-existing beliefs shape the way in which relevance is detected, and the way in which a system detects relevance influences how beliefs are supported or formed.

Forthcoming in *Philosophical Psychology* – Please cite published version

Italics mine). In this sense, the incapacity to see contrary evidence as relevant with respect to the delusional belief can be characterized as a form of resistance to disengagement. This inability to see stimuli as relevant works in two directions: on the one hand, it contributes to the formation of the delusional belief itself by “biasing” the collection of evidence; on the other hand, it reinforces the already-endorsed delusional belief by excluding stimuli that may facilitate its revision. For example, in a case of Capgras the subject falls prey to a pervasive feeling of estrangement towards her spouse. This prompts her to focus on a number of minute details that acquire particular significance: the spouse’s slightly different facial expression, the way in which he ties his shoes, how he uses some words more than others (see Christodoulou 1977). All these details contribute to corroborate the feeling while also paving the way for the endorsement of the delusional belief – i.e. “My husband has been replaced by an impostor”. Once the belief has been endorsed, the patient would then disregard all the signals supporting alternative explanations implying that the person has not been replaced.

The function of *updating* is also extremely important to measure relevance detection: indeed, it comprises “monitoring and encoding incoming information for *relevance* to the task” as well as “revising the items held in working memory by replacing old, *no longer relevant* information with newer, *more relevant* information” (Miyake *et al.*, 2000, p. 57. Italics mine). The connection with working memory is important because the ability to update information seems to require some sort of “temporal tagging” that allows a system to pry apart newer from older data. Interestingly, one of the most common disruptions in delusional (especially schizophrenic) narratives concerns temporal ordering. As Gallagher (2007) observes, people with schizophrenia often experience “difficulties in indexing events in time” and their stories are typically characterized by “a derailing of thought, constant tangents [...] or the compression of a temporally extended story in a single gesture” (p. 218). Besides specific problems with working

memory, patients with schizophrenia also exhibit more general difficulties with the function of updating, such as metacognitive deficits: in particular, they have trouble picking up contextual cues that influence action selection, and often appear unable to monitor their own errors. For example, Proust (2013) discusses a series of cases where patients have trouble correcting motor errors and showing awareness with respect to their own actions: some believe that their limbs are controlled by someone else (delusion of control), while others deny the existence of parts of their body, the need to eat or sleep, or the very fact that they have a body (Cotard delusion) – see p. 299.⁸

The function of *inhibition* can also be meaningfully connected to relevance detection: indeed, this consists in the ability to “deliberately inhibit dominant, automatic or prepotent responses when necessary” (Miyake *et al.* 2000, p. 58). For example, in the first phase of the Stroop test subjects are asked to read a sequence of words naming colors independently of the color of the ink – i.e. they would have to read ‘green’ even if the word is written in blue. In the second phase, subjects are required to read the color of the letters independently of the written word – i.e. if the word ‘green’ is written in blue, they would have to say ‘blue’ and not ‘green’. Results consistently show that participants find the first phase easier than the second, so the latter is normally used to measure a subject’s degree of impulsivity or lack of inhibition. Indeed, the test requires the ability to suppress an automatic response (read the word regardless of the color) in order to focus on a stimulus that is more relevant for the task at hand (read the color regardless of the word). This difficulty to inhibit automatic – although unwarranted – responses may

⁸ One must use caution in citing studies on schizophrenic patients *in general* to support conclusions on people with delusions, since the two groups do not entirely overlap. However, due to our current classification system – i.e. DSM-5 – the presence of positive symptoms such as delusions or hallucinations is still considered necessary for a diagnosis of schizophrenia (see Tandon *et al.* 2013). As a consequence, it is reasonable to infer that the vast majority of individuals currently diagnosed with schizophrenia – and thus included in the studies – would exhibit one or more positive symptoms. Moreover, the authors discussed here are explicitly interested in the subset of schizophrenic patients affected by delusions (see Proust 2013; Gallagher 2007).

Forthcoming in *Philosophical Psychology* – Please cite published version

translate into the delusional subject's inability to suppress one irrelevant explanation (e.g. "My wife has been replaced by an impostor"; "My evil neighbor made my cookies disappear") to focus on more likely ones (e.g. "She must be angry for something I did"; "I must have forgotten to pack them with the rest of the shopping"). Notably, some recent studies found that patients affected by delusional disorder fare significantly worse than controls in the Stroop test, displaying both difficulties in the inhibition of inappropriate responses and inability to "change their attention voluntarily from one aspect of the stimulus to another" (Ibanez Casas *et al.*, 2013, p. 6).

To sum up: the abilities encompassed by the notion of *executive functions* nicely capture various facets of relevance detection. Indeed, shifting requires the ability to disengage from irrelevant tasks in order to engage in others that are more in tune with the system's current goals. Updating allows a system to incorporate relevant information while removing non-relevant ones, while inhibition helps to filter out inappropriate responses. In the next section, I use these empirical measures of relevance detection to characterize the pathological nature of delusions more precisely.

2.3. Cognitive Impairment or Hypervigilance?

Some recent studies have explored the degree of executive impairment in patients affected by schizophrenia and delusional disorder. When flexibility, impulsivity and updating are measured through standardized tests – e.g. Wisconsin Sorting Card Task (WSCT) and Tower of London⁹ – subjects affected by delusional disorder tend to fare significantly worse compared to

⁹ In the WSCT, participants have to match a number of cards with four stimulus cards representing figures of a certain shape, color and quantity. Participants are not told how to match the cards, but only whether a particular match is right or wrong; notably, the required strategy switches during the test – e.g. from matching figures by shape

controls (see Ibanez Casas *et al.*, 2013) and to other psychiatric patients (see Grover *et al.*, 2011). Yet – interestingly enough – delusional subjects have also been found to score *better* than controls in some selected areas: for example they committed fewer errors in the WSCT, thereby suggesting some kind of “hyper-vigilance to selectively abstracted stimuli” (Ibanez Casas *et al.* 2013, p. 7).

This result is significant for two reasons. First, it indicates that the ability to detect relevance in delusional patients is not merely impaired or lacking, but rather *dysfunctional* in a specific sense. Indeed, in the study conducted by Ibanez Casas and her collaborators, patients affected by delusional disorders tended to regard some stimuli as *more* relevant with respect to controls, giving rise to some sort of “hyper-attentiveness” (p. 6). This particular form of hyper-vigilance is in keeping with the phenomenology and narratives of delusional subjects and well reflects important differences between them and other psychiatric patients. For instance, depressed subjects often experience a lack of meaning (e.g. “I felt very still and empty, the way the eye of a tornado must feel, moving dully along in the middle of the surrounding hullabaloo” – Plath 1963, p. 3) whereas delusional patients often report a peculiar *keenness* as well as the feeling of “seeing” connections between things. In particular, in the pre-psychotic stage delusional subjects experience a phase of heightened awareness and emotionality where specific percepts or ideas acquire exaggerated importance. In a first person account of his illness (schizophrenia with delusions of grandiosity), Aaron Reina (2009) reports that he gradually became convinced that one could determine “a person’s feelings, thoughts, and even their personality by the color combinations of their clothing” (p. 3). This heightened attention to colors quickly turned into an obsession: “I believed that these powers were some type of magic [...] I

to matching them by color. In the Tower of London, participants are asked to reproduce a certain arrangement of beads – or disks – in a limited number of moves.

had begun to *see hidden meanings in everything*, from graffiti to architecture and to everyday speech” (Ibid., p. 4. Italics mine). While Reina’s delusions grow more complex, this feeling of enhanced salience becomes more haunting: “From this point on, every bump on the wall of my apartment and every cry in the distance were directed toward me” (p. 5). Second, the fact that delusional subjects do not perform poorly across the board speaks against the objection according to which these patients would exhibit a more general form of *cognitive impairment*. That is, one might object that impaired executive function is not particularly relevant to understanding delusions, because people with schizophrenia exhibit cognitive impairments across most domains. On the contrary, there is evidence in support of the idea that schizophrenic patients perform better than controls in some logical (Owen, Cutting & David, 2007) and probabilistic reasoning tasks (Kasanova *et al.*, 2011). Thus, the results of the tests discussed above combined with the subjective experiences of aberrant salience point towards a more precise clinical picture, where delusional subjects do not exhibit a generic cognitive impairment but rather a more specific disturbance involving executive abilities.

The evidence discussed in this section suggests that delusional subjects exhibit some form of executive dysfunction, which reveals a difficulty in detecting relevance properly in the environment.¹⁰ In the following section I turn to examine the crucial role of *emotions* in the experience and reasoning of delusional subjects. More specifically, I argue that a peculiar interaction between emotional and executive dysfunctions is needed for a subject to fully qualify as delusional.

¹⁰ In this paper I mostly focus on cases where relevance detection disturbances involve external stimuli, such as other people (e.g. Capgras, persecutory delusions), objects (e.g. delusion of reference), or bodily parts (e.g. Cotard). Yet, the account could also be applied to cases involving internally generated stimuli: for example, delusions of thought insertion may be linked to finding certain episodes of inner speech abnormally salient or emotionally upsetting.

3. Delusions as Executive and Emotional Dysfunctions

Drawing a connection between the nature of delusions and the difficulty to detect relevance raises the question of whether such an executive dysfunction suffices to label a subject as delusional. Indeed, other groups of patients – such as the ones exhibiting frontal lobe lesions – tend to perform poorly in tests assessing executive capacities and fluid intelligence (see Roca *et al.*, 2010). However, an important distinction can be drawn between these two groups. While the executive impairment exhibited by frontal patients is usually the consequence of traumatic brain injury, delusional subjects display independent *emotional disturbances* that interact with the executive domain in a peculiar way. In what follows, I argue that it is the influence of emotions that leads to the kind of relevance detection dysfunctions associated with delusions. This allows me to make the pathological nature of delusions more precise, by qualifying them as both executive *and* emotional dysfunctions.¹¹ Moreover, this makes it possible to meaningfully compare delusional subjects to various non-delusional populations, by specifying the way in which emotions affect executive abilities.

3.1. Relevance Detection and Emotions

In the past two decades, the connection between emotions and relevance detection has grown stronger thanks to compelling evidence coming from cognitive neuroscience. Groundbreaking work on the role played by emotions in reasoning has been conducted by

¹¹ The emphasis on the role played by emotions distinguishes the present account from others recently defended in the literature. For example, Gerrans (2014) characterizes delusions as disturbances in the “salience system” responsible for allocating processing resources to information (pp. 61-65). On Gerrans’ view, what makes delusions pathological is that low-level processes – i.e. default cognitive processing – cease to be supervised by high-level ones – i.e. decontextualized processing. This seems compatible with the idea that delusions are executive dysfunctions originated by a subject’s inability to detect relevance in her environment. However, while I argue that such inability is grounded in an emotional disturbance, Gerrans sees the affective component of delusions as a byproduct of the fact that default cognitive processes are unsupervised by decontextualized ones (see p. 221).

Forthcoming in *Philosophical Psychology* – Please cite published version

Damasio through the *somatic-marker hypothesis* (1996). According to Damasio specific areas of the brain – in particular the Ventromedial Prefrontal Cortex (vmPFC) – are associated with the ability to link facts that compose a given situation to the emotion previously paired with that situation in a subject’s experience. These emotional signals or “markers” allow us to activate the appropriate emotions connected to specific situations, playing a crucial role in decision-making and action-guiding processes. Damasio also stresses the evolutionary advantages of this ability, probably originating from a more basic mechanism able to discriminate good or bad responses in terms of survival and later extended to more abstract situations. In healthy subjects, somatic markers guide action by highlighting some components of the context over others, drawing attention to those aspects that are more significant for coping successfully with a certain problem or situation. On the contrary, patients with some form of vmPFC damage are unable to pair emotions with context appropriately and this results in “random and impulsive decision making” (Ibid., p. 1417).

More recently, specific studies on the link between *executive functions* and *emotions* have been carried out by Pessoa and his collaborators (see Pessoa 2013 & 2009; Lim, Padmala & Pessoa, 2009). As opposed to the traditional tendency to label brain regions as either affective (e.g. amygdala) or cognitive (e.g. vmPFC), these studies stress the continuous and complex integration among different systems working together for a common goal. In particular, they show that amygdala and prefrontal cortex cooperate in a number of tasks connected to information gathering and salience detection (e.g. discriminating between threatening and neutral facial expressions). Pessoa also introduces the key notion of *executive competition*, according to which the subcomponents of shifting, inhibition and updating are always interacting in a limited resources scenario. This implies that – within executive functions – “resources devoted to one component will detract from those available to other components” (Pessoa 2009, p. 160). This

Forthcoming in *Philosophical Psychology* – Please cite published version

hypothesis carries important implications for our purposes, because it provides a more specific mechanism through which emotional processing can interfere with executive functions. Indeed, Pessoa's experiments have shown that those elements of the context that are high in emotional significance (e.g. threatening facial expression) take more time to be processed by the subjects and often end up interfering with task performance. Interestingly, this effect is stronger for highly anxious individuals, whose performance seems to decrease exactly because of their enhanced sensitivity towards threatening stimuli that are *irrelevant* to the task. In this sense, emotional content can influence executive functioning in a way that is detrimental for task performance (Pessoa 2009, pp. 164-166). These experiments are important because they show how lower-level phenomena – e.g. immediate affective reaction to a facial expression – can have an impact on relevance detection and higher cognitive states. Indeed, when a stimulus is high in threat (e.g. scared face) attention and control are immediately mobilized with a marked influence of behavior – e.g. processing time increases, performance on other tasks decreases (Ibid., pp. 160-162).¹²

In the next section, I apply the executive competition model to delusional subjects and I show that it sheds light on the different way in which executive functions and emotions interact in pathological and ordinary cases.

3.2. *Application to Delusional Subjects*

Once the connection between emotional processing and executive functions has been established, a key question still needs to be answered: What makes such a relationship distinctively pathological in delusional subjects? The studies discussed above suggest that – in

¹² More evidence in this direction comes from the studies on prosopagnosia, where patients show no explicit recognition of faces but still exhibit a significant affective reaction – measured via skin conductance response – when seeing well-liked family members (see Gabriel, Klein & McCall 2008).

Forthcoming in *Philosophical Psychology* – Please cite published version

many cases – emotions work as *cues* that direct our attention towards some elements of the context while others are filtered out. In particular, emotions have the effect of highlighting some sensory representations over others, so that the former receive prioritized attention in the completion of a task – e.g. when scared faces “pop out” due to their particular emotional salience. Now, in ordinary (non-pathological) cases, there seems to be a certain degree of “fit” between the items highlighted by emotions and the subject’s conscious interests, preferences and goals. For example, a hiker who deeply enjoys solitude would feel happiness while sitting by the side of a lonely mountain lake (see Price 2013, p. 9). On the contrary, in delusional cases this attribution of emotional significance appears importantly misplaced: seeing that a particular lamp-post is unlit, hearing a dog barking, noticing that one’s partner ties his shoes differently are causes of deep emotional turmoil. As a result, the patients’ attention becomes disproportionately directed towards these items that need to be processed to provide a satisfactory explanation of their significance. Thus, the endorsement of a particular persecutory delusion – e.g. “I live in a reality show” – may result from the attempt to explain why other people’s gazes appear particularly invasive or threatening (see Reina 2009).

Taking a deeper look at the notion of *executive competition* proposed by Pessoa (2013 & 2009) allows us to clarify how emotions and executive functions may interact in these situations, making it possible to draw some meaningful distinctions between delusional and non-delusional cases. As I discuss above, Pessoa stresses how emotional contents may modulate executive control in a way that can enhance or impair task performance. On his view, two key factors importantly affect performance outcome: a) the level of *strength* (or arousal) of the stimulus; b) the degree of *task relevance*, that indicates how closely related the stimulus is with respect to the subject’s current goal. Roughly put, when the affective significance of a stimulus is low performance generally increases because executive resources can be “mobilized in service of

Forthcoming in *Philosophical Psychology* – Please cite published version

handling the task at hand” (Pessoa 2013, p. 167). On the contrary, items exhibiting high affective significance – especially when task-irrelevant – tend to compromise performance because they act as “negative distractors” that drain common pool resources from the main task (Ibid., p. 172).

If this account is correct, the executive competition model may provide a more refined explanation of the dysfunctions discussed in §2.3. Indeed, delusional subjects may be continually “distracted” by environmental stimuli that they perceive as threatening (or otherwise negative), and this would cause their common-pool resources to be directed away from those functions needed to complete the task at hand. To make the point more explicit: the executive deficits exhibited by delusional patients – including their proverbial fixity and resistance to contrary evidence – may result from an emotional dysfunction that prompts them to invest a disproportionate amount of resources in processing highly arousing, but task-irrelevant stimuli. This may be what happens – for example – in cases of Capgras delusion: the subject experiences an anomalous affective response to some environmental stimuli (e.g. the spouse’s eye color, the way in which he ties his shoes), which then acquire high strength despite being irrelevant to the subject’s conscious goals and preferences. Given the emotional significance that these details acquire, the subject’s resources are absorbed in attempting to formulate an explanation for the anomalous affective experience. Yet, since emotional and executive resources belong to a “common pool”, and the items processed are both affectively significant and task-irrelevant, we can expect the subject’s executive performance to decline. In particular, it might become difficult for her to filter out automatic but irrelevant explanations (inhibition), as well as to monitor her own errors (updating) and revise her beliefs (shifting). As a consequence, the endorsement of a delusional belief – e.g. “My husband has been replaced by an impostor” – may become more

likely or even inevitable.¹³ In cases like this, the perception of *negative distractors* – i.e. affectively strong, task-irrelevant stimuli – represents the first step on the pathway from emotional disturbance to delusional belief.¹⁴ In this sense, my proposal may be seen as a variation on two-factor accounts of delusions (see Davies *et al.* 2001), where one factor in the subject’s experience – i.e. emotional dysfunction – prevents executive functioning from operating as it should, thereby giving rise to a further disruption – i.e. inability to detect relevance appropriately.¹⁵

Now, how does this pathway diverge from non-pathological cases? Again, the two notions introduced by Pessoa – i.e. strength and task-relevance – allow us to draw some important distinctions. Indeed, many ordinary cases are characterized by a subject’s interaction with stimuli that are *affectively strong* and *task-relevant*. For example, an artist who is passionate about her job may come across a piece of marble that looks perfect for her next sculpture; or, a parent may detect a car driving beyond the speed limit before crossing the street with his children. In such situations, we expect the subject’s performance in the task at hand – i.e. sculpting, street-crossing – to increase significantly, as emotions mobilize the executive resources to direct them toward the

¹³ Are there independent reasons to believe that delusional patients would exhibit such a peculiar emotional dysfunction? Some recent studies have suggested that the emotional disturbances detected in schizophrenia may originate from an aberrant amygdala-prefrontal integration (see Anticevic *et al.*, 2012). The neuro-imaging results are still controversial, but it is worth noting that schizophrenic patients tend to exhibit an exaggerated response to neutral events, suggesting “aberrant responses to environmentally non-salient events” (Anticevic *et al.*, 2010, p. 617). This findings are in line with the pioneering work carried out by Kapur (2003 & 2004), who suggests that a dopaminergic imbalance could be responsible for the peculiar experiences reported by the patients during the psychotic phase. For this reason, dopamine has been at times dubbed “the wind of psychotic fire” and heightened dopamine transmission has been seen as contributing to the formation of meaningful connections between coincident events (Broome *et al.* 2005, p. 26).

¹⁴ Other psychiatric symptoms might arise from the perception of emotionally-laden distractors in one’s environment: for example, phobias and obsessions appear similar to delusions in this respect and may thus share a similar etiological pathway. Similarly, some strong, idiosyncratic religious beliefs may arise from a similar process and thus – in some cases – qualify as delusional (see Graham 2015). I cannot address these issues here, but a closer comparison between these phenomena would be worth exploring.

¹⁵ This point further distinguishes my model from the one defended by Gerrans (2002), who has once proposed a one-stage account for monothematic delusions such as Cotard. However, recently Gerrans has appeared more sympathetic towards two-factor accounts and might have changed his mind on the matter (see 2014, p. 126).

Forthcoming in *Philosophical Psychology* – Please cite published version

relevant goal. Other ordinary cases are rather characterized by a subject's interaction with stimuli that are *affectively weak* and *task-irrelevant*: for example, someone who does not care about football overhears the Super-Bowl results on TV while he is taking care of some house chores. In such cases, the person's overall performance would probably be unaffected, as the stimulus – despite its irrelevance – is not powerful enough to act as a distractor. Notably, there are also cases in which one's interaction with the environment is characterized by the perception of stimuli that are *affectively weak* despite being *task-relevant*. For example, I might realize that coming late to the office one more time would probably get me fired while still being unable to leave my bed in the morning. These cases are particularly interesting because performance may be impaired – and even pathological – for opposite reasons with respect to delusional cases. Indeed, the relevant stimuli may appear exceptionally weak from an affective perspective, thereby depriving the subject of the sufficient degree of attention and motivation required to complete the task.

Taking stock: the executive competition model helps to flesh out the idea that delusions involve both executive and emotional dysfunctions. Indeed, it bridges the gap between several aspects already discussed. On the one hand, implicit and automatic emotional processes shape the patient's attention and direct it towards specific elements of the context that acquire particular relevance. On the other hand, they monopolize executive functions that end up not being available for other tasks, including the ones required to reject the delusional explanation. Moreover, appealing to the notions of strength and task-relevance makes it possible to distinguish more accurately between delusions and other ordinary as well as pathological cases (e.g. depression).

Conclusions

Forthcoming in *Philosophical Psychology* – Please cite published version

In this paper I have drawn on theoretical and empirical considerations to argue that the pathological nature of delusions rests on an executive dysfunction (i.e. capacity to detect relevance in a context) which is, in turn, rooted in an emotional one (i.e. perception of affectively strong, but task-irrelevant stimuli). This proposal constitutes a response to the challenge raised by Bortolotti (2009), according to which delusions do not differ in kind from other irrational beliefs because in both cases the canonical norms of rationality are violated. Although I agree with Bortolotti that delusions cannot be distinguished from other beliefs in virtue of their irrationality, her account leaves us with a puzzle as to what would make delusions pathological. I have argued that this problem can be solved by fine-tuning the notion of rationality to include lower-level abilities such as *relevance detection* – measured via executive functions – and *emotions*. Indeed, it is not until we appreciate the role of these lower-level processes and their influence on belief formation that we can see what makes delusions likely to lead to that decrease in well-being that is the hallmark of pathology. Moreover, qualifying delusions as executive and emotional dysfunctions sheds light on the way in which subjects may be impaired in the evaluation of competing hypotheses, thereby clarifying why – in some cases – the delusional explanation comes to be regarded as the best (or only) option available. This contributes to substantiate Bortolotti’s claim about the epistemic innocence of some delusions (2015b), by uncovering the mechanism through which alternative, non-delusional explanations become unavailable to the subject. This might not apply to every delusion, but it helps to better understand a number of paradigmatic cases in which these beliefs have a negative impact on well-being. My answer to the question “What makes delusions pathological?” thus improves on Bortolotti’s remarks by showing how delusional beliefs are formed in a different way with respect to ordinary beliefs, and how this different etiology leads to a decrease in well-being.

I have substantiated my thesis in two ways: first, I have shown that delusional subjects tend to perform poorly in executive tasks requiring the discrimination between relevant and irrelevant stimuli as well as in tasks requiring the inhibition of inappropriate responses (§2). Second, I have employed the notion of executive competition to show how executive functions and emotional processing interact and I have sketched a possible pathway from emotional disorder to delusional formation. This has also allowed me to compare delusional cases with non-delusional ones and to point out some important ways in which they differ (§3). Generally speaking, I have suggested that delusions may originate from pervasive emotions that prompt the subject to attend to her environment in a peculiar way. Delusional patients exhibit an emotional disturbance that prompts them to perceive several items in their environment as affectively significant despite their being task-irrelevant – i.e. not in-tune with the subject’s consciously endorsed goals and preferences. Given the emotional significance that these items acquire, the subject feels distressed and attempts to formulate an explanation to fit the anomalous experience. A sudden change of expression on the spouse’s face can be interpreted as a sign that the loved one has been replaced (Capgras); a random mural in the city can be read as a hidden message from a famous person (erotomania); a burst of laughter in a group of strangers can be seen as the proof that everyone makes fun of you (persecutory delusion).¹⁶ This peculiar status of befuddlement and alteration in relevance detection then impacts the executive domain, where the resources employed to process some tasks drain from a common pool and continuously detract energy from one another (see Pessoa 2013). Given that the subject’s resources are absorbed in processing affectively significant and task-irrelevant items, executive performance declines

¹⁶ Clearly, some of these delusions are more circumscribed than others: for example, patients affected by Capgras normally believe that only one person has been replaced, whereas patients affected by persecutory delusions exhibit a more generalized sense of anxiety and fear towards others. I am unable to address the issue here, but it would be interesting to see whether the distinction between circumscribed and generalized delusions may reflect an underlying difference in the degree or intensity of the emotional disturbance.

Forthcoming in *Philosophical Psychology* – Please cite published version

significantly. This in turn prevents the subject from assessing competing explanations adequately, thus making the endorsement of the delusion more likely (or even inevitable).

In conclusion, I have shown that Bortolotti's proposal can be amended by fine-tuning the notion of *rationality* in a way that goes beyond the following of procedural, epistemic and agential norms (1). More specifically, I have argued that the pathological character of delusions can be explained by appealing to an executive dysfunction in the subject's ability to detect *relevance* in the environment (2). However, this dysfunction in detecting relevance cannot be divorced from an underlying *emotional* imbalance that gives rise to a peculiar experience of reality, where the delusional subject regards some elements of the context as puzzling and highly significant (3).

References

- Anticevic, A., Van Snellenberg, J. X., Cohen, R. E., Repovs, G., Dowd, E. C., & Barch, D. M. (2010). Amygdala recruitment in schizophrenia in response to aversive emotional material: a meta-analysis of neuroimaging studies. *Schizophrenia bulletin* 38(3), 608-621.
- Anticevic, A., Repovs, G., & Barch, D. M. (2012). Emotion effects on attention, amygdala activation, and functional connectivity in schizophrenia. *Schizophrenia bulletin* 38(5), 967-980.
- Bortolotti, L.
(2015b). Epistemic benefits of elaborated and systematized delusions in schizophrenia. *British Journal for the Philosophy of Science* 0, 1-22.
(2015a). The epistemic innocence of motivated delusions. *Consciousness and cognition* 33, 490-499.

Forthcoming in *Philosophical Psychology* – Please cite published version

(2012a). In defense of modest doxasticism about delusions. *Neuroethics* 5, 39-53.

(2012b). Précis of *Delusions and Other Irrational Beliefs*. *Neuroethics* 5, 1-4.

(2009). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

- Broome, M. R Woolley, J. B., Tabraham, P., Johns, L. C., Bramon, E., Murray, G. K. & Murray, R. M. (2005). What causes the onset of psychosis? *Schizophrenia Research* 79(1), 23-34.
- Christodoulou, G. N. (1977). The syndrome of Capgras. *The British Journal of Psychiatry*, 130(6), 556-564.
- Damasio, A. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society: Biological Sciences* 351, 1413-20.
- Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, 8(2), 133-158.
- Elgin, C.Z. (2007). Emotion and Understanding. In G. Brun, U. Dogouglu and D. Kunzle (eds.) *Epistemology and Emotions*, London: Ashgate, 33-50.
- Freeman, D., Garety, P. A., Kuipers, E., Fowler, D., & Bebbington, P. E. (2002). A Cognitive model of persecutory delusions. *British Journal of Clinical Psychology* 41, 331–347.
- Gabriel, R. H., Klein, S. B., & McCall, C. (2008). Affective reactions to facial identity in a prosopagnosic patient. *Cognition and Emotion* 22(5), 977-983.
- Garety, P., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. . (2001). A Cognitive Model of the Positive Symptoms of Psychosis. *Psychological Medicine* 31, 189–195.
- Gallagher, S. (2007). Pathologies in narrative structures. *Royal Institute of Philosophy Supplement* 60, 203–24.
- Gerrans, P.

Forthcoming in *Philosophical Psychology* – Please cite published version

(2014). *The Measure of Madness. Philosophy of Mind, Cognitive Neuroscience and Delusional Thought*. MIT Press.

(2002). A one-stage explanation of the Cotard delusion. *Philosophy, Psychiatry, & Psychology*, 9(1), 47-53.

- Graham, G. (2015). *The Abraham Dilemma: A Divine Delusion*. OUP Oxford.
- Grover, S., Nehra, R., Bhateja, G., Kulhara, P. & Kumar, S. (2011) A comparative study of cognitive deficits in patients with delusional disorder and paranoid schizophrenia. *Industrial Psychiatry Journal* 20(2), 107–14.
- Hall, J., Whalley, H. C., McKirdy, J. W., Romaniuk, L., McGonigle, D., McIntosh, A. M. & Lawrie, S. M. (2008). Overactivation of fear systems to neutral faces in schizophrenia. *Biological Psychiatry* 64(1), 70-73.
- Ibanez Casas, I., De Portugal, E., Gonzalez, N., McKenney, K. A., Haro, J. M., Usall, J., Perez-Garcia, M. & Cervilla, J. A. (2013). Deficits in executive and memory processes in delusional disorder: A case-control study. *PLOS One* 8(7), 1–8.
- Kapur, S.
(2004). Psychosis as a state of aberrant Salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry* 160, 13-23.
(2003). How antipsychotics become anti-‘psychotic’ – from dopamine to salience to psychosis. *Trends in Pharmacological Sciences* 25, 402-406.
- Kasanova, Z. *et al.* (2011). Optimizing vs. matching: response strategy in a probabilistic learning task is associated with negative symptoms of schizophrenia. *Schizophrenia research* 127(1), 215-222.

Forthcoming in *Philosophical Psychology* – Please cite published version

- Lim, S. L., Padmala, S. & Pessoa, L. (2009) Segregating the significant from the mundane on a moment-to-moment basis via direct and indirect amygdala contributions. *Proceedings of the National Academy of Sciences* 106(39), 16841–46.
- Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of philosophy and psychology* 6(4), 547-569.
- Miyake, A., Friedman, N., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. (2000) The unity and diversity of executive functions and their contribution to complex “frontal lobe” tasks: a latent variable analysis. *Cognitive Psychology* 41, 49–100.
- Owen, G. S., Cutting, J., & David, A. S. (2007). Are people with schizophrenia more logical than healthy volunteers? *The British Journal of Psychiatry* 191(5), 453-454.
- Pessoa, L.
(2013). *The cognitive-emotional brain: From interactions to integration*. MIT Press.
(2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences* 13(4), 160-166.
- Plath, S. (1963). *The Bell Jar*. Faber & Faber.
- Price, C. (2013). The problem of emotional significance. *Acta Analytica* 28(2), 189-206.
- Proust, J. (2013). *The Philosophy of Metacognition. Mental Agency and Self-awareness*. Oxford University Press.
- Reina, A. (2009) The spectrum of sanity and insanity. *Schizophrenia Bulletin* 36 (1), 3-8.
- Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N. & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain* 133(1), 234-247.

Forthcoming in *Philosophical Psychology* – Please cite published version

- Rocca, P., Castagna, F., Marchiaro, L., Rasetti, R., Rivoira, E. & Bogetto, F. (2006) Neuropsychological correlates of reality distortion in schizophrenic patients. *Psychiatry Research* 145(1), 49–60.
- Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neuroscience* 14, 303-316.
- Speechley, W. J., & Whitman, J. C. (2010). The contribution of hypersalience to the “jumping to conclusions” bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience* 35(1), 7-17.
- Surguladze, S. *et al.* (2006). A reversal of the normal pattern of parahippocampal response to neutral and fearful faces is associated with reality distortion in schizophrenia. *Biological psychiatry*, 60(5), 423-431.
- Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R. E., Heckers, S. & Carpenter, W. (2013). Definition and description of schizophrenia in the DSM-5. *Schizophrenia Research* 150(1), 3-10.