Lying by Asserting What You Believe is True: A Case of Transparent Delusion

**Abstract**: In this paper, I argue (1) that the contents of some delusions are believed with sufficient confidence; (2) that a delusional subject could have a conscious belief in the content of his delusion (*p*), andconcurrently judge a contradictory content (not-*p*) – his delusion could be *transparent* (Krstić 2020), and (3) that the existence of even one such case reveals a problem with pretty much all existing accounts of lying, since it suggests that one can lie by asserting what one consciously and confidently believes is true, and (4) sincerity, since it suggests that asserting a proposition you believe is true is neither sufficient nor necessary for sincerity. If I am right about (1) and (2), then (3) and (4) follow easily. Therefore, the paper is mainly devoted to an analysis of transparent delusion and defending (1) and (2).

**Keywords**: delusion, belief, credence, sincerity, lying.

# 1. INTRODUCTION

When one thinks about lying, one’s typical first thought is of people who intentionally assert something they believe to be false. This is why pretty much everyone thinks that liars must, with sufficient confidence, believe that what they say is false – call this the *Belief Condition* on lying. St. Augustine (395), St. Aquinas (1273), Kant (1797), Davidson (1997), Faulkner (2007), Vrij (2008), Carson (2010), Saul (2013), Meibauer (2014), Mahon (2016), and Marsili (2019) are just some of the proponents of this view. Another common conception is that a person is sincere if (and only if) she asserts what she believes is true (Austin 1975; Grice 1989; Williams 2002; Owens 2006) or what she thinks she believes (Ridge 2006) – call this the *Belief Account* of sincerity.

I will argue that this approach to lying and sincerity is incorrect: people can lie even by intentionally asserting what they consciously and confidently believe is *true* and they can sincerely assert what they consciously and confidently believe is false. But the paper is not all about lying and sincerity. Rather, it is mainly about the nature of delusion and it is a natural continuation of my analysis of a phenomenon I named *transparent* delusion (Krstić 2020; see below). The argument that asserting what one believes is true could count as lying is based on the case of Peter, who delusionally believes that the person he sees in front of himself is not his wife and who has the right insight into his condition but who, nevertheless, fails to revise his delusion; his delusion is transparent (Krstić 2020). Therefore, Peter, at least for a while, still has high credence in his delusion (e.g., 0.9 on a scale of 0–1) and this delusion coexists with his judgement to the contrary.

This scenario is important because it raises some interesting questions. For example, if it is possible for Peter to briefly judge that the person he sees is not an imposter but rather his wife *and* that he is delusional while nonetheless still delusionally believing with credence 0.9 that this is *not* his wife, is he lying by asserting ‘This is not my wife’? According to the views based on the Belief Condition, Peter is not lying since he asserts what he confidently believes to be true, a necessary condition for lying is not satisfied, but this surely must be wrong. Because Peter judges both that he is delusional and that his delusion is false, it seems irrelevant that he believes the proposition. In fact, given his overall condition, it seems correct to say that Peter would be lying by asserting a proposition he delusionally believes as *true* – since he would be asserting a proposition he correctly judges to be *false* (he judges that he is delusional and that his delusion is false).

Because Peter can lie by intentionally asserting what he consciously believes to be true, the case in which he lies provides a clear counterexample to the Belief Condition of lying, even in its subtler versions that take account of credences (Marsili 2019), and it suggests that our preferred analysis of lying should be judgement-based (e.g., Stokke 2018); lying seems to involve asserting what one *judges* to be false. In addition, because Peter can sincerely assert what he consciously believes is false (since he judges that this proposition is true), the case in which he is sincere shows that the Belief Account of sincerity is false. Therefore, because Peter can lie by asserting what he believes is true and be sincere by asserting what he believes is false, asserting what you believe is true is neither necessary nor sufficient for sincerity.

Contemporary philosophers have already opposed the Belief Condition and the Belief Account – for instance, Chan and Kahane (2011), Stokke (2014, 2018), and Pepp (2018) – but, to my knowledge, no one has argued that one can lie by asserting what one *consciously*, *confidently*, and *de dicto* believes to be true or that one can be sincere by asserting what one *consciously*, *confidently*, and *de dicto* believes to be false. Extant arguments mainly appeal to subconscious or *de re* beliefs. Therefore, this discussion is novel and important.

However, we still need reasons to accept this interpretation of the case briefly sketched out above. How can a person believe that *p* while consciously judging that not-*p*? Also, Peter’s delusion may not be a genuine belief. Therefore, following my earlier analyses (Krstić 2020, 2023a) and improving them, I will defend the view that Peter suffers from a specific thought-evaluation impairment which makes his judging that *p* ineffective with respect to his conscious belief that not-*p*. And, to this analysis, I will add strong reasons to believe that Peter’s thought should count as a fully-fledged belief. Establishing these two theses will allow me to claim that, at least briefly, Peter both *judges* that *p* and counts as consciously and confidently *de dicto* *believing* that not-*p* (his credence is 0.9) rather than *p*. And, as a result, we can say that Peter’s temporally limited situation is such that he can (i) sincerely assert that *p*, which is what he consciously and confidently believes is false, to any addressee, including himself, and that he can (ii) lie by asserting that not-*p*, which is what he consciously and confidently believes is true, to any addressee, including himself. The fact that he can lie to himself by asserting a proposition he believes is true allows us also to think that he can deceive himself (make his credence less accurate) by lying to himself, a possibility I explore in detail in Krstić (2023a).

All in all, in this paper, I defend the following main claims:

1. at least some delusions (like Peter’s) are beliefs – subsection 2.2,
2. some people believe that not-*p* while judging that *p* and that their belief is false –subsection 2.2.3,
3. these people can sincerely assert what they believe is false (*p*) – subsection 2.1,
4. these people lie when they assert what they believe is true (not-*p*) – section 3.

Subsection 2.1 sets out the *Peter* case. Subsections 2.2–2.3 defend the interpretation, according to which Peter consciously and confidently *de dicto* believes only that not-*p* even though he judges that *p* is true. From this, I argue that Peter is sincere by asserting that *p*, a proposition he consciously and confidently *de dicto* believes to be false. Section 3 argues that, by asserting that not-*p*, a proposition he consciously and confidently believes is true, Peter is lying. Therefore, some liars assert what they consciously and confidently believe is true and the speaker’s conscious judgements matter more for sincerity and lying than their beliefs matter. Section 4 concludes the discussion.

# 2. SINCERELY ASSERTING WHAT ONE BELIEVES TO BE FALSE

## *2.1. The Peter Case*

In my first analysis of transparent delusion (Krstić 2020), I discuss the case of Mr F. (documented by Zislin, Kuperman, and Durst 2011), who claimed that he believed that he was Hitler while at the same time giving a sound argument as to why he cannot be Hitler. I named this condition *transparent delusion*, since it is transparent to the sufferers that they are delusional *and* that their delusions are false.[[1]](#footnote-1) To avoid repeating the same argument, I will offer a series of cases based on a documented case of a male patient in the early stage of recovering from his Capgras delusion, a deluded thought that his wife has been replaced by an identical-looking imposter. The patient from the original case is not transparently delusional but, because his testimony is very interesting, I will slightly modify the case to make the symptoms of transparent delusion easily identifiable. Also, I will use Turner and Coltheart’s (2010) two-factor framework of delusion as a reference point, which I will then gradually refine taking into account relevant features of transparent delusion.[[2]](#footnote-2) This approach will help me to flesh out the most important features of the phenomenon and make my argument easier to follow.

Capgras delusion typically develops because the autonomic response to a specific familiar face is just as if the face was of a stranger; the affective reaction is missing. Because affective information is an integral part of the recognition of familiar faces (Tranel et al. 1995; Breen et al. 2000; Ellis and Lewis 2001), the missing affective response is consciously experienced as ‘the feeling of strangeness’ and so people think that the familiar face is not the face of the relevant familiar person but of someone who looks just like them. Here is one illuminating report: ‘Mum’s walked in the room … well, this picture of Mum … and started talking but it was only a picture of her but it didn’t feel like her’ (Turner and Coltheart 2010: 372). A standard two-factor explanation of this condition goes as follows.

The first factor (impaired thought-formation) starts with a relevant neurological deficiency – an impairment in the person’s perceptual, affective, or mnemonic processing (Breen et al. 2000; Ellis and Lewis 2001; Brighetti et al. 2007) – which prompts the delusion and determines its content by generating abnormal data that affects the person’s perceptual experience. The delusion arises when the perceptual cognitive module either explains this abnormal data using Bayesian abductive inference, the so-called ‘explanation’ model (Coltheart, Menzies, and Sutton 2010; McKay 2012), or simply endorses the input that presents the spouse’s identical lookalike, the so-called ‘endorsement’ model (Davies and Coltheart 2000) – (also, see, Bayne and Pacherie 2004; Davies and Egan 2013).

The second factor (impaired thought-evaluation), according to Turner and Coltheart (2010), consists of failures of unconscious and conscious processes that constitute two monitoring frameworks that check occurent thoughts. The *unconscious* checking system (UCS) either ‘tags’ suspect thoughts as requiring extra conscious evaluation, or ‘passes’ them (as correct) thus conferring conviction. If the UCS ‘passes’ the delusion-generating explanation or endorsement of the input, it reaches consciousness as a proper fully-fledged thought that need not be checked for consistency, which then leads to a ‘(preconscious) feeling of rightness’: it does not feel as if one knows the person one sees and this feeling is taken as correct. Because the feeling of rightness accompanies delusions, Turner and Coltheart say that the person’s UCS is impaired. And from the fact that the patients’ *conscious* checking system (CCS) was not initiated – namely, the person is not using her higher-order reasoning abilities to revise her delusion – they infer that CCS is also impaired. However, if some of the patient’s relevant CCS capacities start recovering, the person will begin to evaluate her delusion and therefore question its truth. Consider now this patient (Turner and Coltheart 2010: 371).

*Opening Case*: (italics added):

**Examiner**: What has made you realise that they’re just imaginations? People telling you that?

**Patient**: No *it’s just myself* that’s done it. I’ve started going through it, and seeing what could possibly happen and what couldn’t happen. … Mary couldn’t suddenly disappear from the room, so there must be an explanation for it. So then I try and work out what. ... She [the ‘double’] knows me way back. The lady knows me way back. She could say things that happened 40 years ago, and I wonder where she gets them from. And then *I worked it out* and I’ve wondered if it’s Mary all the time. *It’s nobody else*.

Recovering from a delusion is a very confusing and stressful state and many patients do not realise what is happening to them; therefore, their responses might be misleading. To avoid this concern, let us say that the patient is Peter, a brilliant philosopher and neuroscientist. Let us further plausibly assume that Peter’s delusion is transparent. We get that, as the casesays, Peter concludes that the lady is Mary by making a series of deductive inferences; he ‘workes it out.’ Nonetheless, given the facts about how transparent delusions are produced and maintained (subsection 2.3; Krstić 2020, 2023), because the first factor is still operational and because the thought-evaluation mechanism still passes the delusion as correct, let us say that Peter still has high credence in ‘the lady is not Mary’ proposition, say 0.9 on a scale from 0 to 1. Finally, as a brilliant philosopher-neuroscientist suffering from a condition that allows correct second-order assessment of his pathological belief, Peter correctly understands his overall cognitive situation (*what* he believes and *why* he believes it). With this in mind, let us add some further steps to the dialogue above (let the examiner’s name be C):

*Peter* (*Opening* *Case* continues):

**Peter**: And then *I worked it out* and I’ve wondered if it’s Mary all the time. *It’s nobody else*.

**Examiner**: So, you now believe it’s Mary?

**Peter**: No, no – I honestly can’t say that I do, C. *It’s obviously Mary but I still confidently believe it’s not Mary*! I would say that my credence in ‘This is *not* Mary’ proposition is 0.9, at least.

**Examiner**: You still believe that she is *not* Mary? This surely can’t be if you judge that she *is* Mary.

**Peter**: That’s the point, C. I am saying that, although my knowledge of Mary and the nature of delusions clearly tells me that the lady is undoubtedly Mary and that I’m delusional, the delusion hasn’t lifted – I still confidently believe that she’s not Mary.

**Examiner**: Are you sure Peter? Maybe, you’re just confused. Remember that guy we had last week?

**Peter**: C, *I’d be lying if I told you it isn’t Mary, but this is what I nevertheless believe*! C, my delusion is transparent!

I will argue that, in this situation, (i) Peter sincerely asserts a proposition he consciously *de dicto* believes is false and that (ii) Peter is right when he says that he would be lying by asserting a proposition he consciously *de dicto* believes is true. These are the main features of *Peter*: Peter

1. judges that *p* is true but he, nevertheless,
2. does not believe that *p*, rather he
3. believes that not-*p* (credence 0.9),
4. because he is a brilliant philosopher-neuroscientist, he correctly assesses his overall condition (what he believes, with what confidence, and why he believes it), and
5. the intuition is that Peter is not lying even though he asserts a proposition he consciously, confidently, and *de dicto* believes to be false; rather,
6. the intuition is that he is sincere – since judgement is more important for sincerity than belief.

These features separate *Peter* from other supposed counterexamples to the Belief Account of sincerity since, in these cases, the speakers assert what they unconsciously (Chan and Kahane 2011; Stokke 2014) or *de re* (Pepp 2018) believe to be false. However, my case may raise one concern. My view is that Peter does not lie because he is sincere but one may think that Peter did not lie because he does not have any deceitful intention and this intention is necessary for lying (I thank the reviewer for this excellent observation). This plausible rival interpretation needs to be countered right away.

### 2.1.1 Why Peter is not Lying

The debate on whether all lies are intended to deceive has been rather dynamic in the last two decades and the philosophers involved in it are divided into two camps. In the first camp, the so-called *deceptionists* argue that all lies are intended to deceive, and, in the second camp, *non-deceptionists* argue that some lies are not intended to deceive (names by Mahon 2016). Parsed as a list of necessary and sufficient conditions, the standard deceptionist view says that

**L**: I lie to you if and only if

L1: I *assert* that *p* to you,

L2: I *believe*, with sufficient confidence, that *p* is false (the Belief Condition),[[3]](#footnote-3)

L3: By asserting that *p*, I *intend* to cause you to

1. believe that *p*, or
2. become more confident in *p*, or
3. believe that I believe that *p*, or
4. become more confident in thinking that I believe that *p*.[[4]](#footnote-4)

According to non-deceptionists, L3 is not necessary for lying, but this view is not our concern at the moment. In *Peter*, Peter believes that not-*p* (not Mary) but intends to cause C to believe that *p*, which he believes to be false, since he believes that not-*p* is true. Therefore, Peter countsas intending to deceive, L3 is satisfied, and he thus counts as lying on the standard deceptionist view (L1 and L2 are also satisfied). But, some versions of deceptionism do not give a false positive in this case. One famous deceptionist, Jennifer Lackey (2013, 2019), came up with an argument that not only avoids this consequence but can also be used to challenge my explanation of why Peter does not count as lying.

Lackey defends the view that all lies are intended to deceive in the following way. She first distinguishes *deceitfulness* (intentionally causing a false belief) from *deceptiveness* (intentionally concealing relevant information) and then says that liars must intend to be deceptive. Vitally, deceitfulness does not involve intending to cause a *believed-to-be-false* belief (L3a,d); rather it involves intending to cause a *false* belief, which may involve the liar’s assessment of the reliability of their own beliefs – as in Peter’s case. Therefore, Peter does not lie on her view because, ‘when [he] states … a proposition that [he] believes is false, [his] aim is not to bring about a false belief in [the examiner] or to conceal [his] own beliefs on the matter’ (Lackey 2013: 244); he is neither deceptive nor deceitful.

This correctly classifies *Peter* but at a very high cost: many examples and arguments strongly suggest that intending to deceive is not necessary for lying (e.g., Carson 2006, 2010; Sorensen 2007, 2022; Krstić 2019, 2022, forthcoming; Sneddon 2021; Michaelson and Stokke 2021; Krstić and Wiegmann 2022) and Lackey’s argument has already been directly challenged many times (e.g., Fallis 2015; Krstić 2019). Let us briefly consider a case that nicely shows that the idea that all lies are intended to deceive is very problematic (derived from Krstić 2019).

*Pinartio*: A vicious murderer, Tony, is hiding from the police in Pinocchio’s house. In search of Tony, the police knock on Pinocchio’s door asking whether Tony is hiding in his house. Pinocchio wants to give Tony away but he is afraid that, if he gives any indication of this to Tony, Tony will hurt him. Luckily, Pinocchio knows that the police know that his nose starts to grow at the very instant he forms the intention to lie and that they know that he knows that they know how his nose behaves, etc., and he also knows that Tony does not know anything about this. Therefore, he asserts ‘Tony definitely isn’t in my house’ to the police. Pinocchio does this not because he wants to deceive the police but rather because he wants them to deductively infer that Tony *is* in his house from the behaviour of his nose and his lie. He also intends that they realise that this was his plan all along (he doesn’t want them to think that he is protecting a murderer).

In *Pinartio*, Pinocchio lies to the police intending to thereby make them realize (i) that Tony is in Pinocchio s house, (ii) that Pinocchio is lying, and (iii) that he is lying because he wants them to deductively infer the true proposition from his assertion and the behaviour of his nose. Pinocchio is not trying to cause the police’s credences in any proposition to become less accurate and he is not trying to conceal anything from them; he is not being deceptive. Therefore, the lie was not intended to deceive in any of the senses discussed in the literature. And because they are in a standard context in which people count as asserting what they say, this seems to be a genuine lie (also, see, Krstić 2022, forthcoming). Krstić and Wiegmann (2022) conducted two experiments to test these hypotheses and an overwhelming majority of their participants had the same intuitions: 95% of their participants thought that Pinocchio did not intend to deceive the police in *Pinartio*.

Cases like *Pinartio*, thus, give us very good reasons to reject the view that intending to deceive is necessary for lying and, if so, then the claim that Peter does not lie because he does not intend to be deceptive is much less convincing. A better explanation is that Peter does not lie because he is being as sincere as possible in his position. And we should believe that Peter is sincere because he says during the interview that (a) he does not believe the content of his judgment, that (b) he thinks that he should believe the content of his judgement because his delusion is clearly false and the judgement is correct, but that (c) he still believes the content of his delusion (d) because of his peculiar condition. There simply does not seem to be anything that Peter is possibly trying to hide from C. I now proceed to specify the view that I will defend.

### 2.1.2 My View

My view is neither that Peter will never abandon his belief that the lady is not Mary – though my central case (Krstić 2020) shows that this is possible – nor that he will not occasionally fluctuate between believing and not believing the same proposition. Also, the argument is not that all delusions are beliefs; the focus is on Peter but, by analogy, it applies to other cases of Capgras delusion as well. The claim is that, at the time of sincerely asserting ‘It’s definitely Mary but I believe it’s not Mary,’ Peter consciously and confidently believes that the lady is *not* Mary while judging that she *is* Mary. Accordingly, my main task here is to show

* + 1. that his delusion that *p* is a belief (his credence is high enough),
    2. that judging that not-*p* will not get the belief immediately revised, and
    3. that he does not believe both that *p* and that not-*p*.

I will defend the following aetiology of Peter’s delusion. Despite judging that *p* is true (‘The lady is Mary’), Peter still confidently believes that not-*p* because of four factors. The first two are (1) his abnormal experience – it doesn’t feel like Mary – and (2) the endorsement of the perceptual input as not depicting Mary. These two factors give the delusion its content. What makes the delusional thought effective as a belief is that (3) the unconscious checking system (UCS) did not tag it as suspicious, and that (4) the conscious checking system (CCS), who has detected that the thought is incorrect, cannot revise it. In short, factors (1) and (2) generate the belief by providing the content, factor (3) makes it operational by passing it, while factor (4) is such that Peter cannot revise it even though he consciously judges that he should revise it. This situation is analogous to the case of, Mr F., who delusionally thought that he was Hitler (*inter alia*) while simultaneously correctly judging that he cannot be Hitler – since he is a Jew and Hitler is not – and that he is ‘a very sick man’ (Krstić 2020: §2).

The following mechanism generates Peter’s delusion. Because the input does not incite the affective response cognitively expected for a perception of Mary (step I), the perceptual cognitive module understands it as not depicting Mary (step II); the delusion results from a normal cognitive response to the relevant input. And given that it corresponds to the input, the UCS passes the thought as correct (step III), which then emerges into Peter’s consciousness as a fully-fledged belief (subsection 2.2.1). Finally, due to the relevant thought-evaluation impairment, CCS is unable to revise the belief even though it is capable of detecting that it is false (step IV) (subsection 2.3; Krstić 2020). Therefore, Peter still believes that not-*p* and only that not-*p*; since he cannot simultaneously and consciously believe both that *p* and that not-*p* (subsection 2.3).

These are the relevant steps in the aetiology of the delusion:

1. the input does not incite the affective response cognitively expected for a perception of Mary.
2. (because of 1) the perceptual cognitive module understands it as not depicting Mary.
3. (since it corresponds to the input) the UCS passes the thought as correct; the thought thus becomes a belief.
4. CCS is unable to revise the belief even though it is capable of detecting that it is false.

In subsection 2.2, I develop the argument in detail. I first defend the claim that the thought is a belief, which is the first main premise in the argument (steps I–III). Following Krstić (2020), subsection 2.3 defends the claim that one may continue to confidently believe that *p*, and consciously so, while judging that not-*p* (step IV).

## *2.2. Peter’s Delusion is a Belief*

### 2.2.1 The Main Argument

Many philosophers argue that delusions are not beliefs. Rather, they might be imaginings (Currie 2000), faulty perceptual inferences similar to illusions (Hohwy and Rajan 2012), default thoughts that occupy the role of beliefs (Gerrans 2014), or even bimaginations, states ‘in-between’ belief and imagination (Egan 2009). Among other things, the obvious paradoxicality of many delusions, the patients’ seeming disregard for the truth, and the fact that they typically do not act on their delusions suggest that delusions are not beliefs.

If delusions are not beliefs, then the full-belief analyses of lying, according to which people lie by asserting what they believe is false, does not misclassify Peter as lying in the *Peter* case: Peter does not assert a proposition he believes is false; he just thinks he believes that the lady is not Mary. Therefore, we need to establish that Peter’s delusion is a belief and I will argue that, even if delusions are not typically beliefs, we have good reasons to regard Peter’s delusion as a belief.Allow me to first present the main argument and then resolve some anticipated objections.

Shah and Velleman (2005) write that to answer the question ‘whether to believe that *p*’ is to answer the question ‘whether *p*’. Obviously, Peter judges that his delusional thought is false; this is the distinguishing feature of transparent delusion. Nonetheless, there is a sense in which the thought still satisfies their criteria of being truth-directed and truth-regulated. The truth, as the standard of correctness, was applied sub-personally: the truth-tracking mechanism presented in steps I–II (subsection 2.1) was initiated by the perceptual cognitive module. However, for a mental state to count as a belief, on this view, the normative standard of being correct if and only if true must also be applied: the mental state must be truth-regulated. On the Shah-Velleman theory, if one judges that *p* and affirms that *p*, if the affirmation produces a standing representation of *p* as true, then this affirmation typically becomes the belief that *p* (however, see Krstić 2020). This standard, in Peter’s case, is also met sub-personally: the UCS’s passing the thought as correct, step III, is a sub-personal application of the normative standard of correctness.

In brief, I argue that steps I–III (Section 2.1) involve the Shah-Velleman type of doxastic reasoning that generates beliefs, but *on a sub-personal level*. Endorsing the input as veridical, steps I–II, is applying the descriptive standard of correctness of belief while passing the belief as correctly depicting the perception, step III, is applying the normative standard. Therefore, Peter’s delusion satisfies both standards to count as belief.

I will now address the three most pressing concerns left open by this short analysis. The first is, if the thought was passed by the impaired UCS, as Turner and Coltheart (2010) argue, then we may say that the normative standard of being correct if and only if true was not applied correctly – the unimpaired UCS would not have passed it. Therefore, the delusional thought is not a belief. The second concern is that, even if it reached consciousness as a fully-fledged thought, this thought is not a belief because beliefs are personal-level phenomena and some person-level features of a belief are missing. For example, on Shah and Velleman’s (2005) view, Peter should believe the content of his judgement, rather than his delusion. These concerns say the thought is not a belief because it was not formed correctly. The final objection is that, even if the thought reached consciousness as a fully-fledged belief, Peter’s judgement will override the sub-personal mechanism and the belief will be revised immediately.

I will address each of these concerns in succession, the first two in the next section and the third in Section 2.2.3.

### 2.2.2 The Thought is Truth-Directed And Truth-Regulated

Turner and Coltheart (2010) posited that an impairment of the unconscious checking system (UCS) caused the sub-personally generated thought to emerge in the patient’s consciousness as a fully-fledged delusion – step III. They reason that, had this system been working reliably, the thought would have been tagged as suspicious, which would then trigger a sense of doubt or unease about it and thereby alarm the CCS. If this is correct, Peter’s delusion is not a belief because it was not formed correctly: the normative standard was not applied.

However, UCS need not be impaired for a delusion to arise. The fact that delusional patients are not surprised by their passed-as-proper delusions when these emerge in their consciousness indicates that a mismatch between expectations and reality was not detected. Turner and Coltheart (2010: 355) find the lack of surprise and the inappropriate feeling of conviction indicative of the impaired UCS. But this reaction is not at all strange: no one should be surprised by not having an affective reaction to perceiving an unknown person that looks identical to their spouse – since this is what Peter and other Capgras patients genuinely perceive.

I do not deny that Capgras patients should be surprised that what they perceive is not their spouse (or some other familiar person) but rather a person that looks exactly like the spouse (the spouse’s ‘picture’). Considering the low probability of an event such as seeing the ‘picture’ rather than the spouse, their conscious checking system (CCS) should signal that this thought should be reconsidered (see, Parrott 2016). The lack of *this kind* *of* surprise indicates only that the patients’ conscious ability to make plausibility judgements is impaired. I also do not deny that these patients should feel surprised by the fact that everybody else recognizes the ‘picture’ as their spouse. This is also a failure of the CCS.

What I do deny is that these people should be a UCS-kind-of surprised by their delusional thought and their not having an affective reaction to the person they perceive: they see the same *face* but not the same *person*. Capgras patients are not aware of the lack of autonomic response to the person they perceive (Coltheart 2005); rather, they just have the experience of seeing the ‘picture’. Therefore, they cannot be surprised by the lack of autonomic response – perceiving the ‘picture’ is not supposed to have one. Allow me to explain.

Perception of familiar people is based on the process of systematically matching inputs with stored representations of familiar items acquired through perception (Ellis and Lewis 2001; Friston 2002a, 2002b; Friston and Stephan 2007). These representations act as top-down predictions of future inputs based on what we know about the world. And matching perception with predictions of familiar people is highly dependent on expected affective reactions (Tranel et al. 1995; Breen et al. 2000; Ellis and Lewis 2001), as opposed to hair colour, hairstyle, a type of garment, or similar. Changes in the latter may even go unnoticed. Therefore, when a vital ingredient such as the affective reaction is missing, the correct perceptual decision is ‘This is not my spouse’, and when the hairstyle is different, the correct perceptual decision is ‘This is not how I remember my spouse’. Because the input does not match the vital ingredient of the spouse-prediction, Capgras patients do not perceive their spouses but rather their ‘pictures’.

Our sub-personal cognitive abilities are based on the work of domain-specific peripheral cognitive modules (Evans 2003). The perceptive cognitive module operates with quite limited information and the UCS checks whether a specific module has performed its function successfully (Turner and Coltheart 2010: 357–362). The UCS has access to some higher-order analytic processes that are built into cognitive top-down predictions; however, it does not have access to *all* of them, and we should assume that the UCS does not have access to those processes that assess epistemic possibilities of things and events – since these require access to working memory and background knowledge and thus belong to higher-order analytic reasoning processes (Evans 2003).

Thinking that the UCS should, on the count of low probability, tag the perception of the ‘picture’ as suspicious entails that it will also tag the perception of an old friend whom you have not seen in ages as suspicious, which is incorrect. Therefore, because Capgras patients do not perceive some familiar people but rather their ‘pictures’, the UCS should not find anything surprising in the delusional thought. Thus, the thought is rational considering the perceptual input (there is no fault in steps I–II), the normative standard of belief-formation was correctly applied by the UCS, and the second factor in Capgras delusion may consist solely of an impaired CCS.

The claim is not that delusional people cannot initially experience a momentofsurprise characteristic to that of seeing a person looking just like someone you know. They may initially be surprised *to* see the ‘picture’ (these are cognitive expectations) but they will not be surprised *that* they see the ‘picture’ (this is veridicality of perception). They may exhibit a kind of (CCS) surprise identical to being surprised to see an old friend in a coffee shop, but they will not exhibit a kind of (UCS) surprise identical to doubting that one is genuinely seeing that old friend (see, Friston and Stephan 2007). In sum, having no UCS-based reason to doubt their senses, Capgras patients need not be surprised by their perception or the resulting delusion although, at first, they may be (CCS) surprised to see the ‘pictures’.

Be that as it may, the second concern is that, because Peter has disregarded consciously accessible information relevant to *whether* *the lady is Mary* due to the impaired CCS, this thought is not sufficiently truth-directed and, thus, it is not a belief even if Peter thinks that it is. This objection assumes that some CCS functions, while not constitutive of perception, are constitutive of the processes whereby we form perceptive beliefs and that, therefore, they are integral to how beliefs are truth-regulated. However, this hypothesis is implausible: because consciously monitoring perceptive information requires working memory resources, this would lead to resource depletion (Gilbert et al. 1990: 610; see Gilbert et al. 1993). Because having your CCS participating in the formation of perceptual beliefs would be evolutionarily disadvantageous, its work is not integral to the formation of perceptual beliefs. What monitors this process is our UCS. Therefore, Peter’s delusional thought may count as a belief.

This concludes the first stage of the argument that Peter’s delusion is a fully-fledged belief. I defended this view in two steps. In the first step, I argued that Peter’s mental state was truth-directed and truth-regulated sub-personally and thus formed in the way beliefs are formed. In the second step, I argued that UCS was not impaired, that it passed the thought because it was consistent with the perceptive input (the first concern resolved), and that CCS need not participate in the formation of perceptual beliefs (the second concern resolved).

In the second stage, I will address the third significant concern, according to which the belief was quickly abandoned as a result of Peter ‘working it out’ that the lady is Mary. I will not deny that Peter may eventually abandon his delusion, nor that he may occasionally fluctuate between believing that *p* and that not-*p*; rather, following my earlier account of the relevant CCS failure (Krstić 2020), I will maintain that judging that *p* (Mary) does not entail that the belief that not-*p* will be abandoned immediately. I will conclude that Peter believes only that the lady is not Mary and that this belief is conscious and *de dicto*.

### 2.2.3. On Revising the Belief Immediately

Subsection 2.2.2 argues that Peter’s delusional belief is inconsistent with some of his other beliefs but that it does not lack the relevant domain-specific consistency (Peter genuinely does not perceive his wife) and that, therefore, only a CCS failure is necessary for his delusion to arise (the second factor: impaired thought-evaluation). In this subsection, I resolve the third concern (the belief should be revised immediately) by offering an account of the relevant second-factor failure responsible for the aetiology and maintenance of Peter’s delusion, step IV in my analysis (Subsection 2.1).

The account of the relevant thought-evaluation failure (second factor) goes as follows (Krstić 2020: §4.1; 2023a: §3.2). Because our thoughts are formed in contexts and because they are connected to certain narratives, they acquire relevant context- or narrative-sensitive associations. In order to subject a particular thought to reality testing (i.e., seeing what’s possible from various perspectives), the thought must first be decontextualized, namely, it must be purged of contextual associations. When this happens, an organism can effectively reality test the thought (i.e., apply various CCS reasoning processes to it) without these associations interfering. The process of thought-evaluation, thus, can be understood as having two steps. In the first step, call it thought-decontextualization, the thought gets purged of its contextual and narrative associations and then, in the second step, it gets analysed through different domains of reality, call it reality-testing, which requires the application of various reasoning processes (both higher-order and lover-order). In the reality test (step 2), the outputs of relevant heuristic (UCS) and analytic(CCS) reasoning processes are *put together* and *applied* to the thought, which is then *evaluated* in light of this information (rendered through different domains of reality).

The above sketch (from Krstić 2020: §4.2) suggests that it is not only that various UCS and CCS abilities must be individually operational for a delusion to be revised, as how Turner and Coltheart (2010) suggest, but it is also that (a) their outputs must be *used* in the reality-test evaluation of the thought and that (b) the thought itself must be decontextualized before the reality test (contextual and narrative associations interfere with the process). Therefore, contraTurner and Coltheart (2010), I argue that the fact that a delusional thought was not revised does not entail that UCS or CCS processes are not working. Rather, it may be that the faulty thought-revision process passed the thought as correct because (a) some functioning CCS abilities were not used in the reality-testing process (a step 2 error), or because (b) the thought was not successfully purged of its contingent cues (a step 1 error) (Krstić 2020: §4.1).

I do not argue that this is the full description of the second factor involved in the aetiology of delusion but only that this particular failure explains transparent delusion. Typically, delusions persist because some analytic reasoning processes are impaired or because they are not purged from contextual and emotional cues. However, people like Peter recognize that their thoughts are false and they do want to abandon them in the face of the evidence and their judgement; the processes constitutive of both steps in thought-revision have been initiated. Therefore, we cannot explain transparent delusion by appealing to these two hypotheses. The remaining hypothesis is that something in the reality testing went wrong: while the contextual cues have been purged successfully (step 1) and the reality test has been initiated (step 2), *the reality test was incomplete*, ‘those outputs of the analytic reasoning processes that signal that the thought is false have been left out when the thought was analysed through different domains of reality’ (Krstić 2020: 194). Because of this step 2 failure (the second factor), the thought was passed as correct.

The explanation, then, is this: prompted by his conscious judgement, (1) Peter’s right hemisphere initiates the revision of his delusion but (2) the process is incomplete because (3) the very thing that signals that the thought is false (the conclusion of the judgement) has been left out of the reality testing (that particular domain has not been included in the test). Because of this failure, (4) the decontextualized reality test ‘passes’ the delusional belief as correct; the conclusion of the judgement is not used to revise the belief. The output of the CCS) ability to judge consciously has been left out of the delusion’s reality test because the right hemisphere is impaired and it cannot conduct the decontextualized reality test thoroughly (see, e.g., Gerrans 2014: §4.6).

I already replied to some objections to this theory that I could anticipate (Krstić 2020, 2023a) and so I will not do this here. This discussion exploits my main idea that ‘*judging that p does not entail acquiring the belief that* *p* or even unbelieving that not-*p*’ (Krstić 2020: 194). This claim is based on the view that the primary function of decontextualized reality testing is connecting various thoughts and reasoning abilities in one simultaneous process; however, if a faulty reality test fails to put some thoughts together, they exist simultaneously but separately. Applied to *Peter*, because the judgement that *p* and the belief that not-*p* were not connected in decontextualized reality testing, Peter’s judgement did not override the delusional belief.

And we do have real-life examples of people who consistently fail to revise their delusions according to their judgements. Mr S., for example, admitted that it is not merely unusual, but unbelievable and virtually impossible that he has two families (he could not believe that he believes that) and yet he could not utilise this awareness to revise this delusion (Alexander, Stuss, and Benson 1979). Judging that he cannot have two families did not trigger a revision of the delusion or an acquisition of the correct belief. Also, in my central case, Mr F. thought that he is Hitler, Satan, David, and Moses knowing fully well that he is none of them and that he was ‘a very sick man’ (Zislin, Kuperman, and Durst 2011). Akratic belief is an example of a non-pathological thought-evaluation failure of this kind and I discuss other cases of non-pathological transparently false beliefs (Krstić 2020).

This particular thought-evaluation failure does not entail that transparent delusions are not beliefs. Doxasticists about delusions think that cases such as superstition – in which people consciously refuse to integrate their superstitious beliefs with their other beliefs – suggestthat good integration with the person’s overall network of beliefs is not a feature essentially constitutive of beliefs and that, therefore, this is not a reason to think delusions are not beliefs (e.g., Bortolotti 2010: 85–88, 93). This view can be resisted (Van Leeuwen 2014) but this does not matter because Peter’s delusion is not analogous to superstitions. Unlike superstitions, Peter’s delusion is integrated, just not fully integrated and it is revisable (the thought-revision has been initiated); it is just not revised *yet*. Therefore, even those who believe that religious and superstitious credences are not beliefs may have good reasons to take Peter’s transparent delusion as a belief.

In conclusion, the third concern (Peter’s delusional belief that not-*p* will be revised as soon as he judges that *p*) is resolved by positing that the incomplete decontextualized reality testing may be the full description of the second-factor failure. Peter’s CCS is recovering and he is now capable of conducting relevant analytic processes, such as judging based on available evidence. Nevertheless, he is still unable to apply the result of his ‘working it out’ to his belief by initiating a thorough decontextualized revision and, thus, the delusional belief will not be revised immediately. Rather, it will continue to be operational despite Peter’s attempts to revise it until his ability to incorporate newly gained insights into the belief’s decontextualized reality testing sufficiently recovers. Peter’s credence in not-*p* (‘the lady is not Mary’) may decrease but, if it does, it will decrease only slightly, which is why he correctly and sincerely asserts ‘It’s definitely Mary but I believe it’s not Mary’.

Importantly, it does not follow from all of this that Peter also acquired the belief that the lady is Mary (the content of the judgement). When one holds inconsistent beliefs, those beliefs are suitably separated into different belief-compartments (see, Egan 2008; Quilty-Dunn and Mandelbaum 2018; Bendaña and Mandelbaum 2021). Peter’s condition cannot be explained in this way since his judgement that *p* and the belief that not-*p* occur in his working memory simultaneously and in the same context, and he is aware of them both. In this situation, the belief that not-*p* prevents the judgement that *p* from generating the belief that *p*. Peter judges that *p* while believing only that not-*p*.

The claim is not that Peter will not occasionally fluctuate between believing and not believing that *p*. The claim is that Peter’s correct judgement (*p*) will not immediately and permanently suspend the belief that not-*p* (not Mary) and, since he asserts that *p* immediately after the judgement, he sincerely asserts what he consciously, confidently, and *de dicto* believes to be false. Finally, because Peter already has a full, confident belief that not-*p*, he cannot be ‘in-between’ believing that *p* (Schwitzgebel 2002, 2010).

This concludes the discussion on the nature of Peter’s delusional thought. I now proceed to draw some valuable insights from his peculiar condition.

# 3. LYING AND INSINCERITY

The *Peter* case invalidates the Belief Account of sincerity: Peter asserts what he consciously and confidently *de dicto* believes to be false (‘This is Mary’) but, because he (correctly) judges both that the proposition he asserts is true and that his belief is false (because he is delusional), he is not being insincere. Rather, he is being as informative as possible during the interview. Therefore, the case suggests that the person’s conscious judgements are more important for determining insincerity than their conscious beliefs. Asserting what one consciously and confidently *de dicto* believes is false is not sufficient for insincerity.

Even more, Peter can lie by asserting that the lady is *not* Mary, which is what he consciously and confidently *de dicto* believes is true, in any context and to any addressee, including himself. Asserting what one consciously and confidently *de dicto* believes is false is not even necessary for insincerity. Consider this variation of *Peter* (see also Krstić 2023a, the *IRS* case).

*Peter Breaking Bad*:

Suppose that Peter hates Mary whom he had married only for her money and that, one night he comes to a conclusion that the lady *is* Mary while nevertheless confidently (credence 0.9) believing that she is not; Capgras patients often live with the ‘imposters’. Knowing that Capgras patients may become extremely violent towards ‘imposters’, Peter quickly decides to use his condition to kill Mary and get away with it. He asks her to make him a sandwich and, once she picks up a knife, he shoots her in the chest. Upon their arrival, Peter tells the police that he has killed an armed burglar disguised as his wife. When they tell him that the lady he has killed must have been his wife, he replies ‘No, that lady definitely is *not* Mary’. Peter asserts this not because this is what he confidently believes to be true but rather because he wants to get away with the murder ofhis wife, Mary.

These are the main features of *Peter Breaking Bad*:

1. Peter judges (correctly) that *p*, but
2. he believes that not-*p* (credence 0.9), and
3. he correctly judges that he believes that not-*p*,
4. he asserts that not-*p* not because he believes it but rather because
5. he wants to cause the police to falsely believe that not-*p* or that he did not know what he was doing (L3a or L3c are satisfied, see subsection 2.1.1),
6. The intuition is that Peter is lying.

The Belief Condition of lying (L2) is not satisfied, Peter does not assert a proposition he believes is false, and thus Peter is not lying according to theories of lying that take the Belief Condition to be a necessary condition for lying. However, the intuition seems clear: even Peter knows that he is lying and that he has killed his wife; he judges that his delusional belief is false and that he has killed his wife. Indeed, by asserting that the lady is not Mary, he can lie to the police, his friends, even to himself (should he suddenly be struck by guilt) – and this lie may even be intended to self-deceive (see, Krstić 2023a). Peter is lying not just despite the fact that the Belief Condition of lying (L2) is not satisfied but rather *because* it is not satisfied: by asserting what he believes to be false, Peter would assert what he judges to be true and he would knowingly admit killing his wife.

We cannot say that Peter lies because he intends to deceive; L3 is satisfied – L3a or L3c in particular. Pretty much all analyses of lying say that asserting what you believe is false (the Belief Condition; L2) *is necessary* for lying. According to these definitions, then, even if L3 is satisfied, Peter is *not* lying because L2 is not satisfied. But, Peter asserts what he believes is true because he intends to lie to the police.

The most immediate concern here is that the case is not relevant since the agent is delusional and his belief is pathological. However, the same kind of thought-evaluation failure exists in the general population; it is just that the contrast between the judgement and the belief is standardly not as shocking and as obvious as in cases such as *Peter* and these situations then go unnoticed. This is my main reason for focusing on transparent delusion. Even though it is bizarre, this condition does appear to be real. Also, patients with pathological disorders are known to lie about their condition to avoid stigmatisation by their environment or to avoid taking medication. If we are genuinely doing a conceptual analysis of lying, we should be able to explain this kind of behaviour. And there is a theory that can be modified to explain the *Peter* cases.

Andreas Stokke (2018: 193, 178) characterizes lying as requiring a conscious intention to answer a question under discussion with a proposition the negation of which the speaker mentally assents to – where this assent is characterized as a conscious action of judging. In short, I lie (if and) only if I assert that *p* while I consciously judge (or am disposed to judge) that not-*p* (see, Pepp 2018) and not-*p* answers the question under discussion.

Stokke’s analysis is designed to deal with cases of *unconscious* beliefs (or false beliefs about what one believes) conflicting with *conscious* judgements that generated conscious beliefs. The person, that is, simultaneously believes that *p* consciously and that not-*p* unconsciously, and the intuition is that she lies if she asserts that not-*p* even though this is what she unconsciously believes. Peter asserts what he consciously and confidently believes is false *and does not believe to be true in any sense*, but this does not entail that the assent/dissent analysis fails. We still should hold that liars must assert propositions they judge to be false (i.e., they dissent to), since this is what explains why we think that Peter is lying. We just need to acknowledge that conscious assent could conflict with a conscious belief. Therefore, the *Peter* cases suggest that our preferred analysis of lying should be judgement-based and that a sensible modification of Stokke’s analysis works pretty well.

# 4. CONCLUDING REMARKS

By analysing *Peter*, the case of a brilliant but delusional philosopher-neuroscientist, I argued that some delusions are beliefs. From this, I argued that asserting what you confidently believe is false (even if you intend your hearer to believe your assertion) need not count as lying. By analysing *Peter Breaking Bad*, I argued that there are situations in which you may count as lying by asserting what you confidently believe is true. My analysis allows us to draw the following conclusions.

First, we have good reasons to think that Capgras delusions are typically beliefs (given that the UCS has no reason to doubt the perceptual input and that CCS does not participate in the formation of these beliefs) and that Peter’s delusion is a belief. Also, *Peter* is a counterexample to the Belief Account of sincerity: Peter sincerely asserts what he consciously, confidently, and *de dicto* believes is false. Therefore, the case supports the view that conscious mental assent (judgement) is more important for sincerity than conscious belief, which offers a nice refinement to Stokke’s (2014) analysis of insincerity. Even though he asserts a proposition he consciously believes is false, Peter is not lying in *Peter* and he does not seem to be insincere either.

*Peter Breaking Bad* is a novel counterexample to the Belief Condition of lying since Peter lies by asserting what he consciously and confidently *de dicto* believes to be true but correctly judges to be false. Peter utters ‘No, that lady definitely is *not* Mary’ intending to lie; this is an essential part of his plan. Peter knows that he is lying and he even intends to cause his audience to believe as true a proposition he judges is false; he intends to deceive by lying. Therefore, the case suggests that the idea that liars must assert what they believe is false (with sufficient confidence) should be abandoned. A slightly modified version of Stokke’s (2018) judgement-based analysis of lying is superior to its rivals. This should be our preferred analysis of lying.

All in all, according to my analysis, asserting what you believe is false is neither necessary for insincerity (*Peter Breaking Bad*) nor sufficient for insincerity (*Peter*) and asserting what you believe is true is neither necessary for sincerity nor sufficient for sincerity (*Peter Breaking Bad*). Some further insights follow from this. For example, in *Peter*, Peter intends to make C more confident in a proposition Peter confidently believes is false (the lady is Mary) but, in doing this, he is trying to inform C, not to misinform him. Therefore, the case indicates that intending to make your hearer believe as true the proposition you believe is false need not count as intending to deceive the hearer. And *Peter Breaking Bad* indicates thatintending to make your hearer believe as true the proposition you believe is false is not even necessary for intending to deceive the hearer: Peter is trying to deceive the police by causing them to believe a proposition he confidently believes to be true.

All in all, these are claims that follow from my analysis:

1. at least some delusions are beliefs and Capgras delusions seem to be normally beliefs,
2. asserting what you believe is false is neither sufficient for lying (even if you intend to cause the audience to believe you) nor for insincerity,
3. asserting what you believe is false is neither necessary for lying nor for insincerity: people can lie or be insincere by asserting what they believe is true,
4. asserting what you believe is true is neither necessary nor sufficient for sincerity,
5. intending to cause you to believe as true what I consciously, confidently, *de dicto* believe is false is neither necessary nor sufficient for intending to deceive
6. [you may intend to deceive yourself by lying, insofar as you lie by asserting a proposition you believe is true but judge is false; see Krstić 2023a].

All of these conclusions are very important and some are quite radical but they are worthy of our attention.

**Acknowledgement**: I would like to thank John Bishop, Neil Levy, the Editor and especially the anonymous reviewer of this journal. I would also like to thank everyone who anonymously commented on this paper in good faith.

# REFERENCES

Alexander, M.P., D.T. Stuss, & D.F. Benson. (1979). Capgras syndrome: Areduplicative phenomenon. *Neurology*, 29, 334–339.

Aquinas, T. (1273). *Summa theologica*, (Fathers of the English Dominican Province, Trans.) Amazon Digital Services.

Augustine. (1952/395). *Treatises on various subjects*, in (ed.) R.J. Deferrari, 53–120. New York: Fathers of the Church.

Austin, J. L. (1962). *How to do Things with Words*. Oxford/Clarendon.

Bayne, T. and Pacherie, E. (2004). Bottom-Up or Top-Down? Campbell’s Rationalist Account of Monothematic Delusions. *Philosophy, Psychiatry, & Psychology*, 11, 1–11.

Bendaña, J. and E. Mandelbaum. (2021). The Fragmentation of Belief. In C. Borgoni, D. Kindermann and A. Onofri (eds.), *The Fragmentation of Belief*, 78–107. Oxford: Oxford University Press.

Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs: International Perspectives in Philosophy and Psychiatry series*. Oxford: Oxford University Press.

Breen, N., D. Caine, and M. Coltheart. (2000). Models of face recognition and delusional misidentification: A critical review. *Cognitive Neuropsychology*, 17, 55–71.

Brighetti, G., Bonifacci, P., Borlimi, R., and Ottaviani, C. (2007). ‘Far from the heart far from the eye’: Evidence from the Capgras delusion. *Cognitive Neuropsychiatry*, 12, 189–197.

Carson, T.L. 2006. The Definition of Lying. *Noûs*, 40, 284–306.

———. (2010). *Lying and Deception: Theory and Practice*. Oxford: Oxford University Press.

Chan, T., and Kahane, G. (2011). The Trouble with Being Sincere. *Canadian Journal of Philosophy*, 41, 215–234.

Chisholm, R.M., and T.D. Feehan. (1977). The Intent to Deceive. *The Journal of Philosophy* 74: 143–159.

Coltheart, M. (2005). Conscious experience and delusional belief. *Philosophy, Psychiatry & Psychology*, 12, 153–157.

Coltheart, M., P. Menzies, and J. Sutton. (2010). Abductive inference and delusional belief. *Cognitive Neuropsychiatry*, 15, 261–287.

Currie, G. (2000). Imagination, delusion and hallucinations. In *Pathologies of Belief*, eds. M. Coltheart & M. Davies, 167–182. Blackwell.

Davidson, D. (2004/1997). Who is Fooled. *Problems of Rationality*, 213–230. Oxford: Oxford University Press.

Davies, M. and M. Coltheart. (2000). Introduction: Pathologies of Belief. *Mind & Language*, 15, 1–46

Davies, M. and Egan, A. (2013). Delusion, Cognitive Approaches: Bayesian Inference and Compartmentalisation. In *The Oxford Handbook of Philosophy and Psychiatry*, (eds.) K.W.M. Fulford et al., 689–727. Oxford: Oxford University Press.

Egan, A. (2008). Seeing and Believing: Perception, Belief Formation and the Divided Mind. *Philosophical Studies*,140, 47–63.

———. (2009). Imagination, delusion, and self-deception. In *Delusion and Self-Deception: Motivational and Affective Influences on Belief-Formation*, (eds.) T. Bayne & J. Fernández, 263–280. New York, Psychology Press.

Ellis, H.D., and M.B. Lewis. (2001). Capgras delusion: A window on face recognition. *Trends in Cognitive Sciences*, 5, 149–156.

Evans, J. St B.T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*,7, 454–459.

Fallis, D. (2013). Davidson was Almost Right about Lying. *Australasian Journal of Philosophy*, 91, 337–353.

———. (2015). Are Bald-Faced Lies Deceptive After All? *Ratio*, 28, 81–96.

Faulkner, P. (2007). What Is Wrong With Lying? *Philosophy and Phenomenological Research*,75, 535–557.

Friston, K. J. (2002a). Beyond phrenology: What can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience*, 25, 221–250.

———. (2002b). Functional integration and inference in the brain. *Progress in Neurobiology*, 68, 113–143.

Friston, K.J., and Klaas, E.S. (2007). Free-energy and the brain. *Synthese*, 159, 417–458.

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

Gilbert, Da.T., Malone, P.S., Krull, D.S. (1990). Unbelieving the Unbelievable: Some Problems in the Rejection of False Information. *Journal of Personality and Social Psychology*, 59, 601–613.

Gilbert, D. T., Tafadori, R. W., and Malone, P. S. (1993). You can’t not believe everything you read. *Journal of Personality and Social Psychology*, 65, 221–233.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge:Harvard University Press.

Hohwy, J. (2013). Delusions, Illusions and Inference under Uncertainty. *Mind & Language*, 28, 57–71.

Hohwy, J. and V. Rajan. (2012). Delusions as Forensically Disturbing Perceptual Inferences. *Neuroethics*, 5, 5–11.

Kant, I. (1999/1797). On a supposed right to lie from philanthropy, in ed. P. Guyer and A.W. Wood, *Immanuel Kant: Practical philosophy*,611–615. Cambridge: Cambridge University Press.

Krstić, V. (2019). Can you lie without intending to deceive? *Pacific Philosophical Quarterly*, 100, 642–660.

———. (2020). Transparent Delusion. *Review of Philosophy and Psychology*, 11, 183–201.

———. (2022). On the Connection between Lying, Asserting, and Intending to Cause Beliefs. *Inquiry: An Interdisciplinary Journal of Philosophy*, (Online First), 1–21.

———. (Forthcoming) Lying, Tell-Tale Signs, and Intending to Deceive. *Dialectica*.

———. (2023a). Lying to Others, Lying to Yourself, and Literal Self-Deception. *Inquiry: An Interdisciplinary Journal of Philosophy* (Online First): 1–26.

———. (2023b). Lying: Revisiting the Intending to Deceive Condition. *Analysis* (Online First): 1–11

Krstić, V. and A. Wiegmann. (2022). What Common Folks Think About Lying Without the Intent to Deceive. *Erkenntnis*,(Online First), 1–24

Lackey, J. (2013). Lies and deception: An unhappy divorce. *Analysis*, 73, 236–248.

———. (2019). Selfless Assertions. In *The Oxford Handbook of Lying*, ed. Jörg Meibauer, 244–251. Oxford: Oxford University Press.

Mahon, J.E. (2016). The Definition of Lying and Deception. In *The Stanford Encyclopedia of Philosophy*, E.N. Zalta (ed.),

URL = <https://plato.stanford.edu/archives/win2016/entries/lying-definition/>

McKay, R. (2012). Delusional Inference. *Mind and Language*, 27, 330–355.

Marsili, N. 2019. Lying and Certainty. In (ed.) Jörg Meibauer, *The Oxford Handbook of Lying*, 170–182. (Oxford: Oxford University Press).

Meibauer, J. (2014). *Lying and the semantics-pragmatics interface*. Berlin: Mouton de Gruyter.

Michaelson, E. and A. Stokke. (2021). Lying, Deception, and Epistemic Advantage. In *The Routledge Handbook of Social and Political Philosophy of Language*, eds. J. Khoo and R. Sterken, 109–124. Routledge.

Owens, D. (2006). Testimony and Assertion. *Philosophical Studies*, 131, 487–510.

Parrott, M. (2016). Bayesian Models, Delusional Beliefs, and Epistemic Possibilities. *The British Journal for the Philosophy of Science*,67, 271–296.

Pepp, J. (2018). Truth Serum, Liar Serum, and Some Problems about Saying what You Think is False. In (eds.) E. Michaelson and A. Stokke. *Lying: Language, Knowledge, Ethics, and Politics*. Oxford, Oxford University Press.

Quilty-Dunn, J., and Mandelbaum. E. (2018). Against dispositionalism: belief in cognitive science. *Philosophical Studies*, 175, 2353–2372.

Ridge, M. (2006). Sincerity and expressivism. *Philosophical Studies*,131, 478–510.

Saul, J.M. (2012). *Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford: Oxford University Press.

Sneddon, A. (2021). Alternative motivation and lies. *Analysis*, 81, 46–52.

Sorensen, R. (2007). Bald-Faced Lies! Lying without the Intent to Deceive. *Pacific Philosophical Quarterly*, 88, 251–264.

———. (2022). Lie for me: the intent to deceive fails to scale up. *Synthese*, 200, 1–15

Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Noûs*, 36, 249–275.

———. (2010). Acting Contrary to Our Professed Beliefs, or the Gulf Between Occurrent Judgment and Dispositional Belief. *Pacific Philosophical Quarterly*, 91, 531–553.

Shah, N., and Velleman, D.J. (2005). Doxastic Deliberation. *The* *Philosophical Review*, 114, 497–534.

Stokke, A. (2014). Insincerity. *Noûs*, 48, 496–520.

———. (2018). *Lying and Insincerity*. Oxford: Oxford University Press.

Tranel, D., Damasio, H., and Damasio, A. R. (1995). Double dissociation between overt and covert recognition. *Journal of Cognitive Neuroscience*, 7, 425–442.

Turner, M., and Coltheart, M. (2010). Confabulation and Delusion: A Common Monitoring Framework. *Cognitive Neuropsychiatry*, 15, 346–376.

Van Leeuwen, N. (2014). Religious credence is not factual belief. *Cognition*, 133, 698–715.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Hoboken, NJ: John Wiley and Sons.

Williams, B. (2002). *Truth and Truthfulness*. Princeton: Princeton University Press.

Zislin, J., Kuperman, V., Durst, R. (2011). ‘Ego-Dystonic’ Delusions as a Predictor of Dangerous Behavior. *Psychiatric Quarterly*, 82, 113–120.

1. Some delusional subjects realize that they are delusional but think that their delusions are nevertheless true. Their delusions are not transparent. I call them *Catch-22* delusions. [↑](#footnote-ref-1)
2. One-factor theories also exist, but I address the rival views (Krstić 2020) so repeating the argument would unnecessarily add content to this paper. [↑](#footnote-ref-2)
3. According to Carson (2006) and Fallis (2013), for example, the Belief Condition says ‘I do not believe that *p*’. This variation makes no difference for our purposes. Also, some argue that asserting is not necessary for lying but this is also tangential to this argument. [↑](#footnote-ref-3)
4. For more sophisticated analyses of L3, see Chisholm and Feehan (1977) and Krstić (2023b). [↑](#footnote-ref-4)