

HOW TO TRAGICALLY DECEIVE YOURSELF

Jakob Ohlhorst

Volume 13, numéro 2, été 2018

URI : <https://id.erudit.org/iderudit/1059499ar>

DOI : <https://doi.org/10.7202/1059499ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Centre de recherche en éthique (CRÉ)

ISSN

1718-9977 (numérique)

[Découvrir la revue](#)

Citer cet article

Ohlhorst, J. (2018). HOW TO TRAGICALLY DECEIVE YOURSELF. *Les ateliers de l'éthique / The Ethics Forum*, 13(2), 48–69. <https://doi.org/10.7202/1059499ar>

Résumé de l'article

Cet article présente le concept d'auto-illusion tragique. En prenant la notion de base selon laquelle l'auto-illusion consiste en une croyance motivée à l'encontre de meilleures preuves, je soutiens qu'il existe des cas extrêmes d'auto-illusion qui persistent même face à des preuves contradictoires incontestables. J'appelle ces cas : auto-illusion tragique. On pourrait soutenir que des preuves d'une telle force excluent la possibilité d'auto-illusion et qu'il s'agirait plutôt de délire. Afin d'éviter cette conclusion, j'introduis le concept wittgensteinien de certitudes ou de propositions charnières (hinges) : des admissions qui se situent au-delà d'une justification de nature probante. Un type de certitudes en particulier – iHinges, qui sont adoptées pour des raisons d'ordre motivationnel – rend compte du phénomène d'auto-illusion tragique. Ces certitudes justifient que le sujet rejette la preuve sans que cela implique, de son propre point de vue, une perte de rationalité. Subséquemment, je traite de certaines objections qui peuvent être soulevées contre cette explication de l'auto-illusion.



HOW TO TRAGICALLY DECEIVE YOURSELF

JAKOB OHLHORST

GRADUATE STUDENT, UNIVERSITY OF COLOGNE

ABSTRACT:

This paper introduces the concept of tragic self-deception. Taking the basic notion that self-deception is motivated belief against better evidence, I argue that there are extreme cases of self-deception even when the contrary evidence is compelling. These I call cases of tragic self-deception. Such strong evidence could be argued to exclude the possibility of self-deception; it would be a delusion instead. To sidestep this conclusion, I introduce the Wittgensteinian concept of certainties or hinges: acceptances that are beyond evidential justification. One particular type of certainties—iHinges, which are adopted for motivational reasons—explain the phenomenon of tragic self-deception: they warrant the subject's dismissal of the evidence without loss of rationality from the subject's point of view. Subsequently, I deal with some objections that can be raised against this account of self-deception.

RÉSUMÉ :

Cet article présente le concept d'auto-illusion tragique. En prenant la notion de base selon laquelle l'auto-illusion consiste en une croyance motivée à l'encontre de meilleures preuves, je soutiens qu'il existe des cas extrêmes d'auto-illusion qui persistent même face à des preuves contradictoires incontestables. J'appelle ces cas : auto-illusion tragique. On pourrait soutenir que des preuves d'une telle force excluent la possibilité d'auto-illusion et qu'il s'agirait plutôt de délire. Afin d'éviter cette conclusion, j'introduis le concept wittgensteinien de certitudes ou de propositions charnières (hinges) : des admissions qui se situent au-delà d'une justification de nature probante. Un type de certitudes en particulier – iHinges, qui sont adoptées pour des raisons d'ordre motivationnel – rend compte du phénomène d'auto-illusion tragique. Ces certitudes justifient que le sujet rejette la preuve sans que cela implique, de son propre point de vue, une perte de rationalité. Subséquemment, je traite de certaines objections qui peuvent être soulevées contre cette explication de l'auto-illusion.

INTRODUCTION

Self-deception is emotionally motivated belief against better evidence. This is a common way of glossing the phenomenon. It is a very expansive notion: for example, wishful thinking will also fall under this umbrella. There is a debate on whether self-deception needs to be intentional. I ignore this question, because intentional self-deception excludes certain phenomena that I take to be clear cases of self-deception (cf. Mele, 2001, p. 9).

I think that there is a special form of self-deception, which is usually overlooked. I call it *tragic self-deception*: in such cases the subject is able to dismiss *any*, even compelling, evidence. This is the sort of self-deception from which, for example, religious fanatics suffer: the tragically self-deceived are in that state because, to them, certain things *cannot* be true.

Tragic self-deception is overlooked because it does not fit with our usual epistemological concepts. Either the contrary evidence seems to be too weak, so that it leads to ordinary self-deception, or the evidence is considered to be too strong. In the latter case, the subject would have to be mad or delusional, rather than merely self-deceived, in holding such a belief.

Note that “delusional” is ambiguous. There is the medical sense—such a *pathological* delusion means that the subject suffers from some mental illness. I will come back to how this sense of “delusional” appears to exclude tragic self-deception in the case of compelling evidence. The other sense is loose talk: “he is delusional” just means there is something very wrong with his beliefs. This looser sense of “delusional” does not seem to demarcate a class of its own and is often synonymous with “self-deceived.” I shall therefore ignore it (cf. Bortolotti and Mameli, 2012).

Our usual concepts of self-deception provide only a limited arsenal of ways to avoid an abhorred conclusion: selective sampling, selective treatment, biased weighing, and misinterpretation of evidence (Mele, 1997). These self-deceptive tricks cannot dispel compelling evidence against the false belief. Therefore, we overlook the possibility of tragic self-deception. A regularly self-deceived subject confronted with sufficiently strong evidence would therefore snap out of it and find truth. Such a subject who failed to do so would be so deeply irrational that that subject would be pathologically delusional—his or her mind would have to be defective.

Nevertheless, I think we should take the possibility of tragic self-deception seriously. Indeed, if we extend our conceptual framework to include *hinges* or *certainties*, then we can account for tragic self-deception. Hinges were introduced by Ludwig Wittgenstein in *On Certainty* (1969). We can use the notion to account for how and why tragic self-deception arises. Additionally, we can explain away the impression that tragic self-deception is impossible and that it collapses into pathological delusion.

I will begin by presenting an example of what I take to be tragic self-deception. Indeed, I believe that such cases are quite common. I then explain how our understanding of rationality threatens the concept of tragic self-deception. I believe that this idea explains why tragic self-deception seems to collapse into delusion. I will then introduce certainties as the feature of our epistemological apparatus that is able to explain tragic self-deception.

My account of certainties loosely follows Wittgenstein's (1969). A peculiar form of such certainties can cause self-deception. I call them *iHinges*, and they are adopted for subjective motivational reasons. This paper shows how they explain the possibility of tragic self-deception, where the subject is absolutely resistant to contrary evidence.

While this account deals with the mentioned objection of irrationality, other objections can be raised. I conclude by explaining and responding to those other objections: that my definition of tragic self-deception seems to make it indistinguishable from pathological delusion as it is usually defined, that it does not account for stubbornness or the nagging doubts usually accompanying self-deception, and that it is too heavy-handed an approach for something so straightforward as self-deception.

TRAGICALLY DECEIVING ONESELF

Tragic self-deception arises when a subject has a belief that is *immune to any, even compelling, evidence*. This is analogous to regular self-deception being mere resistance against evidence. What is meant by "immunity," and what do I mean by "compelling evidence"?

A belief is immune when there is no possible evidence that could rebut or undermine it. I name this *immunity to evidence* because the belief could be lost due to other psychological factors like being shocked or overwhelmed, but no *evidential* reasons could dislodge it.

Evidence is compelling when it is impossible to ignore, when it entails what it is evidence for, and when the entailment is also impossible to ignore. When neither the evidence nor its evidential relations can be ignored, this makes the evidence compelling: the evidence forces us to believe.

Now this seems to make tragic self-deception an inconsistent concept: how can we be immune to believing something that we're forced to believe? I will work out this inconsistency more precisely in the section on tragic self-deception and madness. Ultimately, I shall argue that the appearance of inconsistency is resolved by epistemic certainties. But first I will give an example of tragic self-deception and develop the notion to show that it is nevertheless rather natural.

MISBEGOTTEN: A young woman is accused of a heinous crime: the evidence laid out by the prosecutor against her is crushing and she has

confessed to it repeatedly. Notwithstanding this, her parents remain convinced of their baby's innocence. If they accepted that their daughter was a criminal, they would be devastated: in their eyes, they did everything for their daughter, they gave her their values, and their identity is built on their idea of their daughter's moral character.¹

This is an example of what I named tragic self-deception—it is the first example that comes to my mind when I am thinking about self-deception in general. It fits the idea that self-deception occurs when people believe something against better evidence because they desire or fear it to be true.

I have to note, however, that I think that self-deception is a multifarious phenomenon—I am not certain that *all* self-deception is a matter of believing against better evidence.² Let us then take a closer look at the example.

The parents of “Misbegotten” are self-deceived. Most people would be called so if they believed something because of some motivation without appropriately accounting for their evidence. They believe that their daughter is innocent because they desire her to be so and because they fear the contrary. Meanwhile, such a desire has no epistemic force. Nevertheless, they dismiss the compelling contrary evidence they have: all that has been laid out in the judiciary proceedings.

Still, “Misbegotten” differs from run-of-the-mill examples used for self-deception as characterized by motivated resistance against evidence. Take the story of Sid and Roz as an example.³ Sid is in love with Roz; Roz, however, does not share this sentiment. Because she fears hurting his feelings and because she nevertheless appreciates him as a friend, Roz does not simply rebuff him. Rather, she implicitly communicates in a way that any sensible person would understand that she is not interested. Unluckily, Sid is not sensible; he is in love. Given his emotional state, he misreads Roz's refusals as encouragements.

Sid has evidence against *p* but because of his motivational state—desiring Roz to love him—he misinterprets the situation and comes to believe that she also loves him. The difference from “Misbegotten” is that, if Roz had explicitly told Sid that she does not want to be with him, then Sid would snap out of it. He is not deluded either, so runs the thought, except in a hyperbolic sense. And he would not resist *any* evidence against the desired outcome: if Roz wrote him a letter, he would maybe want to double-check with her; but he would not out of hand dismiss it as a forgery sent to him by his malevolent competitors. That is, Sid is insensitive but not immune to evidence.

The parents in “Misbegotten”, however, are so deeply entrenched in their belief that no evidence could convince them otherwise. I therefore call “Misbegotten” a case of *tragic self-deception*—tragic, because it is characterized by total immunity to evidence.⁴ Tragic self-deception is then a special kind of self-deception, which is characterized by its being emotionally motivated and immune to evidence—something that Mele's classical model (2001, p. 50-51) arguably cannot account for.

Given the above considerations, I shall follow Alfred Mele's (1983, p. 370; 2001, p. 50-51; 2009, p. 267) deflationist approach to self-deception in order to characterize tragic self-deception. The following characterization is a modification of Mele's:

- S enters *tragic* self-deception in acquiring a belief that p *iff*
- (1) the belief that p that S acquires is false;⁵
 - (2) S treats *the evidence* relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way;
 - (3) this biased treatment is a nondeviant cause of S's acquiring the belief that p; and
 - (4) the body of *evidence* possessed by S at the time *entails not-p*, S is *aware of the entailment*, and *the evidence is such that S cannot ignore it* (cf. Mele, 2009, p. 267).⁶

By *entailment* I do not mean mere material implication but rather strict entailment. It is the fourth condition that makes self-deception tragic, as it strengthens the evidential relation to entailment and makes the evidence impossible to ignore. Consequently, the subject is aware of the evidence and of what it entails.

For the general debate on self-deception, I shall be relying more or less on Mele's position and work (2001). I do not think that tragic self-deception and deflationist self-deception are the only types of the phenomenon. Nor do I think that the intention to deceive oneself is a prerequisite for being self-deceived. For, if we survey how the nonphilosophical literature describes self-deception, we find instances of motivated belief against evidence.⁷ These are philosophically interesting phenomena, whether they are called self-deception or not. But if we consider the sort of examples I have mentioned, these seem (at least to me) well captured by the term "self-deception" or "tragic self-deception."

REMAINING SUFFICIENTLY RATIONAL

Egocentric and Objective Rationality

Tragic self-deception raises a whole nest of issues. Immediately, there is the question of rationality. Holistically speaking, if we take into account all the reasons a subject might have, tragic self-deception may be rational. It allows the subject to pursue his or her life as before. However, I am concerned with the epistemology of self-deception and focus on the self-deceived's epistemic rationality. There are two senses of epistemic rationality: objective and egocentric rationality.

Objectively, the self-deceived are clearly irrational. We do criticize people for having objectively irrational beliefs—for example, when an agent does not account for all the evidence that agent has available or when an agent believes mutually exclusive things. There is then a standard of rationality that everybody can be measured against. Richard Foley calls this "objective rationality" (Foley,

1991, p. 369), and it is circumscribed by how a knowledgeable outside observer would evaluate the situation.

But this is not everything. There also is an inside perspective on rationality, which evaluates the subject's own point of view:

We are sometimes interested in evaluating your decisions from your own egocentric perspective. Our aim is to assess whether or not you have lived up to your own standards ... or, perhaps, the perspective you would have had were you to have been carefully reflective. (Foley, 1991, p. 367–368)

Following Foley, I call this weaker notion *egocentric epistemic rationality*. We all have standards of coherence for ourselves: our beliefs should be compatible with each other and with the evidence we have available. If we become *aware* of incoherence, either between our beliefs and our evidence or amongst our beliefs, then we are required to revise either a belief or our interpretation of the evidence. If we failed to account for all the evidence we are aware of or to deal with our incoherence, we would become egocentrically irrational, believing things we believe to be incompatible.

There is the common idea that we cannot be egocentrically irrational if we want to preserve our sanity. That is, we as mentally healthy subjects are incapable of believing contradictions while being aware of their contradictoriness.⁸ In the same way, clear and immediate evidence against our beliefs cannot be ignored just like that—if we did, something would be wrong with our minds. Becoming consciously egocentrically irrational appears to be a form of madness; it implies a mind in disarray.

Even if a consciously egocentrically irrational person were not crazy, it still would be extraordinary. Could we make sense of what that person was doing and thinking?⁹ If somebody seems to be consciously egocentrically irrational—that is, holding inconsistent beliefs while aware of the inconsistency, not accounting for evidence while recognizing it, but otherwise behaving normally (if that is possible)—then we need some explanation for what is going on. I do not think that a sane subject can be consciously egocentrically irrational.

Tragic Self-Deception or Madness?

So, we are also subject to egocentric rationality. The household of our beliefs is subject to diverse constraints. Anything we are or would, upon reflection, become aware of falling under these constraints constitutes our egocentric rationality. As mentioned, we have to appropriately account for the evidence that we are aware of, and we cannot believe a contradiction while being aware of it; otherwise, we are consciously egocentrically irrational.

Those who are tragically self-deceived seem to be consciously egocentrically irrational—they dismiss evidence that they recognize to be compelling. Conscious egocentric irrationality implies that the subject is mad. If all instances of tragic self-deception are instances of madness, then what good is the concept? It's just madness. This reasoning creates the appearance that cases of tragic self-deception with sane subjects are impossible. Consequently, tragic self-deception does threaten a subject's egocentric epistemic rationality and thereby itself as a concept.¹⁰

In tragic self-deception, the available evidence is compelling. The subject cannot ignore that evidence without becoming consciously egocentrically irrational. But if that subject accounted for the evidence by making the appropriate inferences, this would lead to a contradiction with the strongly confessed self-deceptive beliefs. The subject would again be consciously egocentrically irrational. How could a subject be tragically self-deceived without being egocentrically irrational, given that as long as such a subject holds on to a self-deceptive belief he or she will end up in an apparent incoherence?

Prima facie, there seem to be two options on how to interpret what is going on here: either the subject ends up egocentrically irrational due to the compelling evidence and his or her immunity to it, or the evidence was not actually as compelling as it was made out to be.^{11, 12}

The first option means that the subject is out of his or her right mind. Conscious egocentric irrationality is not possible for a sane subject—consequently, the consciously egocentrically irrational person must suffer from some mental illness. Most probably, that person is delusional. Delusion is defined as “a false belief ... that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary” (DSM-5, 2013, p. 819). Now, if tragic self-deception were nothing but a symptom of mental illness, calling it a kind of “self-deception” would be unwarranted.

The other option is that actually the evidence is not compelling. If that is the case, then this is not a case of *tragic* self-deception. Rather it is regular evidence and the subject is just ordinarily self-deceived.

Notwithstanding this appearance, I want to defend the notion of tragic self-deception as something that occurs with fanatics or in cases like “Misbegotten”. I want to maintain that a sane subject can stay both egocentrically rational and immune to any evidence, even if that evidence is compelling. This is possible because the evidence may be compelling, but it cannot be *certain*. In a way, I concede that compelling evidence is not as compelling as it might seem at first sight. Evidence cannot compel us to just any belief, even when we have accurately assessed it. That would be an unrealistic requirement. Evidence may always be defeated. The evidential relation is one of entailment, and every *modus ponens* has its *modus tollens*. Finally, there is a class of doxastic states that defeat any evidence. These states are called *certainties* or *hinges*.

HINGING

Regularly

I shall argue that tragic self-deception is grounded in a peculiar feature of our epistemologies: *hinges* or *certainties*—two terms that I use interchangeably. Historically speaking, hinges go back to Ludwig Wittgenstein’s *On Certainty* (1969). But they have recently regained popularity in internalist epistemology (cf. Coliva and Moyale-Sharroch, 2016).

Certainties or hinges are acceptances¹³ that are beyond evidential justification. This immunity against evidence stems from their peculiar role: they are so general or fundamental that one cannot possibly find any noncircular evidence for or against them. They are the acceptances that are attacked and rendered visible by sceptical arguments. Wittgenstein gives the following example:

91. If Moore says he knows the earth existed etc., most of us will grant him that it has existed all that time, and also believe him when he says he is convinced of it. Has he also got the right ground for his conviction? (Wittgenstein, 1969, p. 14e)

As Wittgenstein points out, there is noncircular evidence neither *for* nor *against* such propositions. They are beyond the game of giving reasons. Consequently, they are also beyond doubt. For Wittgenstein, genuine doubt is controlled by evidence—we cannot really doubt something that is beyond evidence; it would just be empty gesturing (Wittgenstein, 1969, p. 18e).

These evidentially ungrounded certainties form the “rock-bottom” on which the whole building of our knowing and believing is erected (Wittgenstein, 1969, p. 33e). Wittgenstein uses the bedrock metaphor because of the hinges’ solidity—as mentioned, we cannot really doubt hinges, and we do not make sense if we do because of their foundational role. Accepting certainties is necessary for evidential relations, communicative interaction, and practical projects to work. Without hinges, we would not be able to investigate, tell, or do anything intentionally.

So how do hinges work? Mostly they are hidden and implicit. They form the fixed background in front of which we operate. They are the things we take for granted and on which our epistemic activity turns. But in peculiar situations, they can come to the fore. Consider the following example.

ROSE: As you read this paper, you see how a rose appears out of thin air, falls to the ground, and shatters. The shards dissolve into thin air again. You do not trust your eyes and refuse to believe that a rose really appeared and disappeared just now.

I would like to point out that the evidence in “Rose” is compelling, you *see* the rose appearing—if the evidence is not strong enough for your tastes, strengthen

it as you like. The certainty that is active in this example can be formulated as follows: “midsized massive objects do not just appear and disappear.” This is certain. You would have a hard time furnishing evidence for it. And evidence against it, such as “Rose”, is dismissed even if it is compelling. Indeed, the certainty *warrants* dismissing any evidence against it.

For, if you were to accept the evidence from “Rose”, you would lose the ability to know many things about objects—especially if the rose’s appearance remained a unique or random event. Your memory would become useless; you could not tell others about things elsewhere. You would not even be able to intentionally get milk from the fridge the way you usually do, as it could have disappeared in the meantime. You would only ever be able to intentionally go check whether there is still milk. Depending on your milk management, that might be what you actually do, and, in that case, imagine yourself to be more reliable.

The constancy of objects is not the only certainty. There are large swaths of interconnected propositions that can be taken to be certain—for example, the belief that the world is more or less as we perceive it to be, the idea that causal relations hold between certain events, or the conviction that other living beings also have consciousness. These certainties inform our world picture (Wittgenstein, 1969, p. 24e). We need such certainties, lest we be unable to act intentionally or to believe genuinely. We would not be able to interact with the world in any way. Hinges are a necessary condition for epistemic and practical agency.

To use a slogan: without certainties, mind-to-world, world-to-mind, and mind-to-mind adaptation would break down. That is, neither could the mind adopt a picture of the world nor could it act upon the world or exchange thoughts with other minds. It is therefore both epistemically and practically *rational*¹⁴ to treat certain fundamental beliefs as certainties or hinges. From this particular role flows the epistemic rationality of dismissing any evidence that is incompatible with our hinges.

The attentive Wittgensteinian will already have realized that I do not follow an orthodox interpretation of what hinges are, because I treat them as logically related to other beliefs and maybe even truth apt instead of animal (a rule of our form of life) and beyond evaluation (cf. Moyal-Sharrock, 2016). Indeed, I am not at all certain that this epistemic account of certainties (cf. Kusch, 2016) accounts for Wittgenstein’s own view. Nonetheless, I believe that such an epistemic account of hinges is the way to go; it is a powerful tool. In other words, I am not wedded to the noble enterprise of reconstructing Wittgenstein’s own philosophy.

Variant Hinges

Things are however not as simple as they may have seemed until now: there is no natural class of certainties. Being a hinge is not an essential feature of propositions. Rather, this depends on the environment in which we live and the kinds of organisms that we are: a member of the San people living in the thirteenth

century has and needs a different world picture than an Indian Brahman from the third century, and their world pictures are profoundly different from a dolphin's.

A salient example of how hinges can be under tension or shift is the theory of relativity. Wittgenstein was aware that we changed hinges in that case. Notably, he would refuse to say that people's former certainty that time and space are independent measures was *mistaken*. Rather, they were "confused" about what time and space are—namely, something that, for lack of better words, bends and stretches (Wittgenstein, 1969, p. 39e). Indeed, if you look at how we think, talk, and behave today, we still largely treat space and time as independent. Another example that Wittgenstein raises is that of faith:

Catholics believe as well that in certain circumstances a wafer completely changes its nature, and at the same time that all evidence proves the contrary. And so if Moore said 'I know that this is wine and not blood', Catholics would contradict him. (Wittgenstein, 1969, p. 32e)

These examples show that there is no salient criterion of which propositions should be hinges—it is contingent for any proposition *p* whether it is certain or not. Rather, hinges arise out of how we treat certain propositions. We take them to be certain and indubitable—we take them for granted. Hinges fulfil a certain functional role in the households of our believing and they are hinges because they play this functional role.

Life Is Easier on iHinge

This opens up space for propositions to be treated like certainties that have little to do with grounding our epistemic, communicative, and practical activities. People do not treat only the existence of the world, or the acceptance that other members of their community will understand what they say, as certain. There may be also more subjective emotional or existential hinges.

Tragic self-deception results from this category of subjective hinges. If we compare the examples of the "Rose" and the misbegotten daughter, then the parallels become clear. In both cases, there is apparent and compelling evidence for some proposition *p*. But the agent dismisses this evidence instead of adapting her beliefs to it. She does this because, from her point of view, things just *could not* be as the evidence would indicate—that is, *p*.

Things could not be such because the agent is *certain* of propositions that imply that not-*p*. Therefore, the evidence cannot be accurate. Thus, the agent's hinges are the grounds on which she dismisses the apparent evidence. Interestingly enough, neither with regular hinges nor in tragic self-deception does the certainty need to be held explicitly. The certainty may become apparent only if we asked the subjects why they dismissed their evidence. The mechanism works automatically.

Further, the certainties preserve the agents' egocentric epistemic rationality when they dismiss even compelling evidence that goes against their hinges. Their functional role as certainties make the dismissal of any contrary evidence egocentrically rational. The same also holds for tragic self-deception: the tragically self-deceived may dismiss any evidence going against their fundamental certainties, without becoming egocentrically irrational. Their self-deceiving hinges parasitize the egocentric rationality-preserving function of regular hinges. I shall call the hinges leading to tragic self-deception *iHinges*,¹⁵ given their central role for our self-image.

What makes certainties into *iHinges*? It's the basis on which they are adopted. As mentioned, regular hinges serve to make action, communication, and inquiry as such possible. They are a structural requirement. Without them, we would end up with an infinite regress or a circularity of evidence justifying beliefs about evidence. All our beliefs and actions would lose their rational grounds. Wittgenstein therefore likes to call them "logical" requirements (Wittgenstein, 1969, p. 7e, 20e, 58e).

iHinges do not play the same fundamental role. They are necessary for our lives in another way. Rather than being necessary for our world picture, they are hinges necessary for our self-image. *iHinges* are the response to questions like "What sort of person am I?" and not to questions that are dealt with by other hinges such as "What is a person?"

Take, for example, the conviction that the people close to you genuinely care about you and vice versa. Evidence for this is extraordinarily hard to come by. All of their behaviour is compatible with the fact that, ultimately, it is only their own enjoyment they seek in interaction with you. Indeed, there are people who argue for a crude *Homo oeconomicus* anthropology, claiming that all actions simply result from an egocentric utility calculus.¹⁶

Nevertheless, we trust ourselves and our close friends to spend time with us and support us not because they expect some corresponding return value from this. In Kantian terms, we have a hinge that to our close friends we are an end in ourselves and not a means to something else. Without this certainty, the very notion of friendship would fall apart. Imagine how empty the idea would be. The only way to avoid this hinge is by believing that you have no friends at all.¹⁷

In other words, *iHinges* may help give sense to our lives. They tie together a self-image, helping us to deal with our desires, doubts, and fears. *iHinges* play an existential role. If such a certainty becomes unhinged and is accepted to be false or doubtful, it is not our *very capability* to interact with the world that is at stake. We still have access to all the categories required to act and believe rationally.

Instead, what happens if we lose an *iHinge* is that we fall prey to an existential crisis: many maxims on which we have put our stakes until now, many evaluations that we took to be certain, and many ideas about who we are would fall

apart. While it would not hinder us from doing or investigating things by robbing us of the prerequisite world picture, it would undermine our motivations, values, and ideas, thus throwing us into lethargy.

We accept iHinges to avoid such existential devastation. They are a psychological or motivational rather than an epistemic necessity for living our lives. In sum, a certainty is an iHinge if and only if it has all the structural trappings of a certainty – immunity to evidence, implications for a wide class of other beliefs, and so on – but it is adopted because it is necessary for preserving our motivations and self-image rather than for preserving our epistemic and practical agency as such. One consequence from this is that, from the point of view of objective rationality, our iHinges ought to be susceptible to evidence. The fact that they are not, because of their existential role, generates problems like tragic self-deception.

PICK YOUR IHINGE FOR YOUR TRAGEDY

iHinges lead to the motivated manipulation of evidence in tragic self-deception. Thus, in “Misbegotten”, the parents have invested everything in their daughter: they instilled her with their morals and values, they cared for her, and scolded her when she did something wrong. In sum, they built their lives around their idea of their daughter.¹⁸ This means that, ultimately, their daughter’s moral integrity has become more important to them than their own epistemic access to the world. If they accepted that she did bad things, their world would break down. Much of what they had done and believed since she was born would lose its meaning.

At first sight, as in any case of tragic self-deception, this appears to be egocentrically epistemically irrational. In her parents’ eyes, the young woman could not do any wrong. Notwithstanding this, the evidence is clear and compelling: she *did* do something wrong. But for the parents, their daughter’s innocence is certain. This allows them to dismiss the unpleasant evidence as fabricated or misleading without a loss of egocentric rationality. Through their iHinge, they tragically deceive themselves in the face of compelling evidence.

They do this as follows: They are aware of the evidence laid out by the prosecutor, and they are aware that this evidence entails that their daughter is guilty. It is compelling, and they cannot simply ignore it. But the parents are also certain that their daughter is innocent. Nothing can epistemically trump a certainty; therefore, their iHinge that the daughter is innocent defeats the evidence from their egocentric point of view. Either the iHinge rebuts the evidence by *modus tollens* or the evidential relation is somehow undercut. How the evidence is defeated will depend on the circumstances.

Consequently, from their egocentric point of view they deal with the evidence in an appropriate manner. They dismiss evidence that in their eyes cannot be accurate in favour of a certainty. The certainty even warrants taking some

account of why the evidence is undercut or rebutted to be the most plausible explanation. This allows them to preserve their egocentric epistemic rationality in tragic self-deception and, consequently, to avoid delusion. In short, we are tragically self-deceived when an iHinge is incompatible with the world and there is compelling evidence for this incompatibility.

This, then, is the basic idea of tragic self-deception: certainties that are adopted for motivational reasons lead to dismissal of accurate evidence. I will now respond to four objections that can be raised against the iHinge view of tragic self-deception. The first three have been raised against Mele's deflationist stance but apply to my view, too. The last one specifically attacks the approach of using hinges to account for self-deception.

Nagging Doubts

Some authors argue that self-deception is essentially characterized by a certain inner tension (Audi, 1997, p. 104; Bach, 1997, p. 105; Losonsky, 1997, p. 121). In other words, when we are self-deceived, we will be aware that something is amiss. We will frequently revisit our deceptive belief and have "nagging doubts" (Losonsky, 1997, p. 121) about it. This point has been raised in response to a paper by Alfred Mele, where he argues that self-deception as investigated by empirical psychology arises out of our biases (Mele, 1997, p. 93-95), a view that does not account for these supposedly essential nagging doubts. The idea behind the objection is that the self-deceived subject is somehow implicitly aware of his or her deceit, hence the nagging doubts.

In tragic self-deception, there can be no such nagging doubts at all. Doubt is *excluded* by the very nature of certainty—iHinges need to be beyond doubt to fulfil their role. Nevertheless, a form of tension remains in tragic self-deception, but it is an outer rather than an inner tension. As Bach observes, the self-deceived subject is at stark odds with reality—"truth is dangerously close at hand" (Bach, 1997, p. 105). First, the subject's belief is false; second, contrary evidence is easy to come by. The self-deceived subject will therefore quite frequently come up against such contrary evidence—having to dismiss it *each time*. This is an outer tension.

Consider how someone would react to not only a single, but repeated occurrences of "Rose": suddenly roses start popping up and disappearing all over the place. The first few times, she would probably simply dismiss it; but with time she would start treating the events differently. She would start to ask others whether they had experienced the same thing as she has, she might seek medical help, and she might even try to examine the phenomenon more closely. In other words, she would start to display the same behaviour as someone doubting—mostly her own sanity. She would be subject to outer tension.

Analogously, those who are tragically self-deceived—always coming up against contrary evidence, always having to do the epistemic work of dismissing it—

may at a certain point start to display such doubting behaviour. Maybe they would not doubt their own sanity but rather the sanity of people in their environment. While this phenomenon is not the inner tension that Mele's critics demand, it may still suffice as an explanation for why there are "doubts" accompanying self-deception.¹⁹

Delusion

Like deflationist self-deception, the notion of tragic self-deception raises a further spectre: that tragic self-deception is mere delusion. More specifically, the question is, why would cases of tragic self-deception not simply be cases of *pathological* delusion? The worry also arises for other accounts that take resistance against evidence as a starting point to account for self-deception. This is so because diagnoses for delusion are also characterized by immunity against evidence:

Delusion. A false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. (DSM-5, 2013, p. 819)

I emphasize the pathological side of delusion because in ordinary language someone who is self-deceived may be called deluded without being a case requiring psychiatric treatment. Tragic self-deception skirts the abyss of pathological delusion. It is epistemically so irrational that it has itself the air of being pathological; the tragically self-deceived looks like a case for a therapist—for example, a psychoanalyst. Nevertheless, I think the tragically self-deceived are quite sane—there is no work for the medically trained psychiatrist.

A nice example that demonstrates the fine line between delusion and tragic self-deception is the romantic partner who, out of jealousy, is self-deceived about being cheated on. How exactly does this self-deceived person differ from someone who needs psychiatric help for morbid jealousy? Morbid jealousy is a syndrome characterized by "a range of irrational thoughts and emotions, together with associated unacceptable or extreme behaviour, in which the dominant theme is a preoccupation with a partner's sexual unfaithfulness based on unfounded evidence" (Kingham and Gordon, 2004, p. 207).

Apart from the "unacceptable or extreme behaviour," someone who is tragically self-deceived fits the morbidly jealous's profile. Epistemically speaking, the self-deceived person is indiscernible from the delusional patient in this case. So, is tragic self-deception nevertheless just a form of delusion?

While this is worrying at first sight, we should keep in mind that psychiatry is anything but an exact science.²⁰ Furthermore, delusions do not form an aetiologically unified class of phenomena. This is in contrast to illnesses like schizophrenia or depression, each of which is unified by a range of symptoms and

neurological characteristics. Already morbid jealousy itself is only the surface description of several other illnesses—for example, it occurs frequently in cases of schizophrenia (Kingham and Gordon, 2004, p. 207).

I surmise that, rather than being a phenomenon of the same kind and at the same level of generality as delusion, self-deception plays the role of a symptom *of* delusion. That is, delusions will always be accompanied by behaviour that can be described as deflationary self-deception, but there can also be isolated self-deception that is not a symptom.²¹

This would explain why the two are hard to separate but nevertheless distinct. An additional argument may be, that for someone to be diagnosed as pathologically deluded, their agency and ability to live a normal life would have to be seriously impaired. Arguably, cases of tragic self-deception exactly succeed at preserving agency and a *partially* normal life. We can push this line of thought further: one of the strategies in clinical psychiatry for distinguishing pathological behaviour from merely unusual behaviour is whether it poses a threat to either the subject or the subject's environment.²² The fine line separating delusion from tragic self-deception would then be how much danger the particular case poses.

This differentiation is rather tentative. But given the main goal of this paper—introducing the phenomenon of tragic self-deception—it may be sufficient to show how delusion may tie in with tragic self-deception. Note, however, that, given the multifariousness of delusion, each syndrome must be treated separately, and that the moves we made for morbid jealousy will not necessarily work universally (cf. Mele 2006, p. 116-123).

Stubbornness

A different, though similar, difficulty is how to differentiate tragic self-deception from mere stubbornness about being right. Kevin Lynch (2013) raised the problem and proposed a viable solution in the same go. The problem, as with delusion, is that stubbornness is intuitively distinct from self-deception, even though it is also characterized by motivated immunity against evidence. A stubborn person does not care about the evidence and will stick to his or her guns.

Lynch brings to attention the idea that acceptances can be adopted *because of* emotional attitudes towards their content or they can be adopted *independently* of content for some other reason. According to Lynch, stubbornness is content neutral (Lynch, 2013, p. 1342–1344). In tragic self-deception, the subject cares about what he or she adopts as a certainty—the subject has an emotional attitude towards what he or she accepts, be it fear, desire, or something else. Meanwhile, in stubbornness, as introduced by Lynch, the content of the beliefs adopted is irrelevant. It is about being right in general, not about any particular belief content. In contrast, in tragic self-deception, we are certain about certain propositions.

That is, if Lynchean stubbornness about being right also operates on certainties, then the relevant hinges will not be about some specific content. Rather, they would take the form “Those who do not share my ideas are intellectually careless” or something in this vein. This hinge would then warrant the dismissal of any counterargument.

Tragic self-deception turns on certainties that are adopted because the hinges’ content fulfils a certain motivational, psychological, or existential role for us. This distinguishes it from stubbornness as proposed by Lynch, for which the content of beliefs does not play any role or only a secondary one. Meanwhile, nothing seems to preclude the possibility that stubbornness as introduced by Lynch is a form of self-deception.

Cannons at Sparrows

Another issue may be the hinge approach itself. Using hinges to deal with self-deception may look like shooting cannons at sparrows. Bringing out the heavy apparatus of certainties to deal with something as straightforward as self-deception when we have so many other options may seem overblown. Are we then exaggerating?

I have two things to say about such general doubts: First, if we grant the notion of hinges a role in our epistemology anyway, then the apparatus is already there, and it will play a pervasive role rather than stay confined to some peripheral phenomena. Second, certainty is not as heavy as it might seem: while there are these very fundamental hinges that ground inquiry and agency, there are more everyday, lighter, ways to acquire something similar to certainties. For example, we take things for granted all the time: you do not always check whether your bicycle brakes still work and whether there is enough pressure in your tires—you just hop on (cf. Wright, 2004, p. 190).

Another line of this objection is to argue that we do not need hinges. We have everything we need with the well-researched menagerie of biases that psychology has to offer. So why use another category, certainties? First, biases are only distorting effects. That is, they strengthen or weaken our credences. Arguably, such distortion effects cannot account for the complete dismissal of strong evidence as occurs in tragic cases of self-deception.

Second, even if biases were able to account for the discarding of compelling unavoidable evidence, they do not give the same kind of explanation as hinges do. Biases are psychological mechanisms; they leave open how a subject treats his or her beliefs, epistemically speaking. We are always subject to epistemic pressures and demands. The dismissal of compelling evidence demands some epistemic explanation if anything does. Biases do not go far enough in that job because they underdetermine what happens in tragic self-deception—so we need something else to lift that explanatory weight: certainties.

Another take on this objection is to argue that tragically self-deceived subjects *take their evidence in a way that* will allow them to infer the self-deceptive conclusion. The flippant response to this would be that, if they are able to reinterpret their evidence, then arguably this is not sufficiently compelling evidence for tragic cases. But I do not think that this would hold up. Rather, I would argue that in cases of tragic self-deception, people do not refuse the evidence by pointing to other independent evidence. This simply is not what happens in tragic self-deception.

The parents in “Misbegotten” cannot point to any strong enough positive evidence for why their child would *never* do such a thing, however privileged it may be; no past evidence would be able to trump or rebut the current *compelling* evidence. At least I do not know what such past evidence would have to look like to warrant dismissing the compelling current evidence. Instead, it just could not be otherwise to them—it is inconceivable.

Additionally, self-deceived subjects would read their past evidence differently in the light of an iHinge. As mentioned, mere biased treatment of past evidence without certainties would not suffice to generate tragic cases of self-deception because the counterevidence could not be compelling in such cases. Something needs to explain the strength of the skewed interpretation of the past: iHinges do this job.

A different issue is borderline cases. There are hinges that are adopted as much out of epistemic curiosity as out of personal pride or desire, which makes it hard to determine whether the subject is self-deceived or not. Take for example Albert Einstein’s refusal to accept the indeterminacy of quantum phenomena. That “God does not play dice” is a classic example of a certainty. Now take some fictitious counterpart of his, Twinstein, who stakes a lot of his ego on his supposedly correct insights into the nature of reality. Is Twinstein tragically self-deceived? It is hard to say. However, I do not think we need to worry. As vagueness appears in many mental phenomena, especially defects, it is not as troubling as it may seem at first sight: we just have to live with it.

LOOKING BEYOND

What lessons are there to be drawn from the hinge-account of tragic self-deception? First, it shows that and how one can be tragically self-deceived without being pathologically delusional or egocentrically irrational. Cases where a subject proves to be absolutely resistant to evidence constitute an important subclass of self-deception. Tragic self-deception shows how extreme the phenomenon can get. Additionally, hinges explain how the subject can be in such a state without becoming incoherent or egocentrically irrational—thereby avoiding the open and apparent contradictions that certain patients of delusion run into. Even somebody tragically self-deceived will appear to be more objectively rational than someone with, for example, anosognosia. Anosognosia is one symptom of hemiplegia among others, where patients do not recognize that

half their body is paralyzed. They confabulate in an often-inconsistent manner to explain away their handicap and their incoherent behaviour.

Furthermore, iHinges give us a tool to diagnose what exactly is going on in tragic self-deception. Notably, they give us a structure of beliefs that tragically self-deceived subjects have. This contributes to understanding what is happening in such cases: how subjects start treating certain beliefs as fundamental and thus are led astray. Indeed, iHinges may be at work in an even broader range of cases of self-deception than only the most extreme cases. In the long run, we may also be able to differentiate different types of self-deception according to the epistemic mechanism at play.

Finally, the account of tragic self-deception as a manifestation of iHinges is able to tie in self-deception with internalist epistemology. Given its structure, internalism always is subject to a certain pressure to account for what is going wrong when something goes wrong with our rationality. My account can be used to explain how tragic self-deception is possible for an egocentrically rational subject. From an internalist point of view, the difficulties are especially pressing because tragic self-deception poses a threat to internalist accounts, raising the spectre of egocentric irrationality.

NOTES

- ¹ A similar example is considered in Bortolotti and Mameli (2012) and Murphy (2012).
- ² For example, Pascal's-Wager-style undertakings of bringing oneself to believe something can be argued to be a typical though different instance of self-deception (Jones, 1998, p. 167).
- ³ One of Mele's favourite examples, as the following not necessarily complete list shows (Mele, 1987, p. 125; 2001, p. 26; 2006, p. 110; 2009, p. 262; 2010, p. 746).
- ⁴ It is their fate to believe this. Tragedy is always based on fate.
- ⁵ This is part of Mele's characterization, but I would not strictly exclude the possibility of being self-deceived even while believing the truth.
- ⁶ This characterization deviates from Mele's on several counts (i.e., the *italics*): first, he simply gave sufficient conditions for cases of self-deception, while I give sufficient and necessary conditions for *tragic* self-deception. Second, Mele used "data" instead of "evidence." I am carefully optimistic that experiential data can be adequately captured by some corresponding fine-grained proposition. I will therefore talk about evidence as propositional. Third, the fourth condition is modified so as to make the evidence compelling.
- ⁷ Consider for example, how supporters of US president Donald Trump are occasionally described by the press: because of their disdain for Hillary Clinton and their attachment to their hero, they manage to not recognize his moral faults (e.g., Allen, 2016; Shapiro, 2016; Douthat, 2017).
- ⁸ The impossibility of believing contradictions is not uncontested (see, for example, Priest, 1986, p. 102).
- ⁹ By this *egocentric irrationality*, I do not mean a dialetheist toying around with liar sentences or metaphysical postulates about the nature of god. These are metalinguistic phenomena.
- ¹⁰ Noordhof (2003, p. 83-88) argues for something similar, though on different grounds. If the subject is aware of the contrary evidence, but still maintains his or her belief, then that subject cannot be self-deceived because self-deception essentially involves instability in the face of contrary evidence. His example (Noordhof, 2003, p. 76) is structurally similar to my MISBEGOTTEN, so I would call it another case of tragic self-deception. Meanwhile, Noordhof appeals to the intuition that we cannot be self-deceived if we are aware of the contrary evidence, because there would apparently be no deception. I will not defend my position against this claim, but simply point out that he takes a fine-grained view of what self-deception is, while I take the broader deflationary approach (Mele, 2001) that bases itself on the circumstance that instances of motivated resistance against evidence can be and are described as self-deception.
- ¹¹ A third possibility is that the subject is not immune and loses his or her belief, but then that subject would not be self-deceived anymore.
- ¹² I thank my anonymous reviewers for pressing me on this point.
- ¹³ It is disputed whether hinges are beliefs, as many take beliefs to be essentially guided by evidence. I therefore use the more neutral notion of acceptance.
- ¹⁴ The epistemic rationality in that case is however atypical: it is not evidential but rather consequentialist or, maybe, transcendental.
- ¹⁵ Thanks to Nikolaj Pedersen for this apt nomenclature.
- ¹⁶ This is a hinge in its own right that often cannot to be dislodged by any amount of argument.
- ¹⁷ This hinge is compatible with the idea that friends betray each other. Betrayal presupposes a friendship to be betrayed.
- ¹⁸ Without making their daughter the centre of their life, they would hardly get so far off-track in their self-deception.
- ¹⁹ What I describe here is only a disposition to display doubting behaviour; however, I doubt that one effectively must be under permanent tension in order to count as self-deceived.
- ²⁰ This can, for example, be seen with the constantly changing catalogue of syndromes and with the structure of their definitions: "S suffers from I if she/he fulfils at least x of the following y criteria."

- ²¹ Mele takes a similar strategy, pointing out the multifariousness of morbid jealousy. Either the phenomena of self-deception and delusion belong to entirely different classes or self-deception is a symptom of the delusion (Mele, 2006, p. 121). See also Bortolotti and Mameli (2012).
- ²² Take as an example the phenomenon of senile dementia: there are many patients where no psychiatric treatment is necessary. They live their lives as they've done the past ten, twenty years—with habit protecting them from the dangers of their condition. But at a certain point they may start forgetting more crucial things—e.g., to turn off the stove—or begin to wander aimlessly and get lost, thereby beginning to pose a threat.

REFERENCES

Allen, Cynthia M., "Self-Deception is the Key to Justifying a Vote for Trump," *Fort Worth Star-Telegram*, August 4, 2016.

Alston, William P., "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives*, vol. 2, 1988, p. 257-299.

American Psychiatric Association (ed.), *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, 5th edition, Washington, D.C., American Psychiatric Publishing, 2013, p. 819.

Audi, Robert, "Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele," *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 104.

Bach, Kent, "Thinking and Believing in Self-Deception," *Behavioral and Brain Sciences*, vol. 20, 1997, p. 105.

Bortolotti, Lisa, "Delusion," in N. Zalta, Edward (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2016 edition, 2013, URL: <https://plato.stanford.edu/archives/spr2016/entries/delusion/>.

Bortolotti, Lisa, and Matteo Mameli, "Self-Deception, Delusion and the Boundaries of Folk Psychology," *Humana.mente*, vol. 5, no. 20, 2012, p. 203-221.

Coliva, Annalisa, and Danièle Moyal-Sharrock (eds.), *Hinge Epistemology*, Leiden, Brill, 2016.

Douthat, Ross, "The Bannon Revolution," *The New York Times*, October 17, 2017, URL: <https://www.nytimes.com/2017/10/11/opinion/steve-bannon-revolution.html>.

Foley, Robert, "Rationality , Belief and Commitment," *Synthese*, vol. 89, no. 3, 1991, p. 365-392.

Jones, Ward E., "Religious Conversion, Self-Deception, and Pascal's Wager," *Journal of the History of Philosophy*, vol. 36, no. 2, 1998, p. 167-188.

Kingham, Michael, and Harvey Gordon, "Aspects of Morbid Jealousy," *Advances in Psychiatric Treatment*, vol. 10, no. 3, 2004, p. 207-215.

Kusch, Martin, "Wittgenstein on Mathematics and Certainties," in Annalisa Coliva and Danièle Moyal-Sharrock (eds.), *Hinge Epistemology*, Leiden, Brill, 2016, p. 48-71.

Losonsky, Michael, "Self-Deceivers' Intentions and Possessions," *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 121-122.

Lynch, Kevin, "Self-Deception and Stubborn Belief," *Erkenntnis*, vol. 78, no. 6, 2013, p. 1337-1345.

Mele, Alfred, "Self-Deception," *Philosophical Quarterly*, vol. 33, no. 133, 1983, p. 365-377.

———, *Irrationality: An Essay on Akrasia, Self-Deception, Self-Control*, Oxford, Oxford University Press, 1987.

———, "Real Self-Deception," *Behavioral and Brain Sciences*, vol. 20, 1997, p. 91-136.

———, *Self-Deception Unmasked*, Princeton, Princeton University Press, 2001.

———, “Self-Deception and Delusions,” *European Journal of Analytic Philosophy*, vol. 2, no. 1, 2006, p. 109-124.

———, “Have I Unmasked Self-Deception or Am I Self-Deceived?” in Martin, Clancy (ed.), *The Philosophy of Deception*, Oxford, Oxford University Press, 2009, p. 260-276.

———, “Approaching Self-Deception: How Robert Audi and I Part Company,” *Consciousness and Cognition*, vol. 19, no. 3, 2010, p. 745-750.

Moyal-Sharrock, Danièle, “The Animal in Epistemology: Wittgenstein’s Enactivist Solution to the Problem of Regress,” in Coliva, Annalisa and Danièle Moyal-Sharrock (eds.), *Hinge Epistemology*, Leiden, Brill, 2016, p. 24-48.

Murphy, Dominic, “The Folk Epistemology of Delusions,” *Neuroethics*, vol. 5, no. 1, 2012, p. 19-22.

Noordhof, Paul, “Self-Deception, Interpretation and Consciousness,” *Philosophy and Phenomenological Research*, vol. 67, no. 1, 2003, p. 75-100.

Priest, Graham, “Contradiction, Belief and Rationality,” *Proceedings of the Aristotelian Society*, vol. 86, 1986, p. 99-116.

Shapiro, Ben, “You Can’t Pretend Trump’s Flaws Away,” *National Review*, October, 2016, URL: <http://www.nationalreview.com/article/440742/donald-trump-supporters-self-delusion>.

Wittgenstein, Ludwig, *Über Gewissheit = On Certainty*, Oxford, Blackwell, 1969.

Wright, Crispin, “Warrant for Nothing (and Foundations for Free)?” *Aristotelian Society Supplementary Volume*, vol. 78, no. 1, 2004, p. 167-212.