



This is a repository copy of *Self-deception about truthfulness*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/177843/>

Version: Accepted Version

Article:

Sleat, M. orcid.org/0000-0003-2518-6940 (2022) Self-deception about truthfulness. *European Journal of Philosophy*, 30 (2). pp. 693-708. ISSN 0966-8373

<https://doi.org/10.1111/ejop.12711>

© 2021 John Wiley & Sons Ltd. This is the peer reviewed version of the following article: Sleat, M. (2021). Self-deception about truthfulness. *European Journal of Philosophy*, 1–16., which has been published in final form at <https://doi.org/10.1111/ejop.12711>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley’s version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Self-Deception about Truthfulness

Of the many things which we hold beliefs about, one will, unless we lack any concern for it at all, relate to the question of our own truthfulness: the extent to which we believe we are truthful in how we come to form, maintain, and revise our beliefs. The disquieting notion which this paper seeks to explore is that it is not only possible for us to be self-deceived as to our own truthfulness, itself not an unfamiliar possibility, but that there is good reason to suspect certain aspects of the way we understand and value truthfulness, especially as it features in political life, make it something which we may be particularly prone to being self-deceived about.

The significance of this possibility in an important sense stands free of any particular social or political context; it would be an important thing to know about ourselves if true. It does, however, have increased salience in our current political climate where truthfulness has become deeply embroiled in the major conflicts and tensions currently playing themselves out within many liberal democratic societies. Some of this has been directed at revisiting the perennial question of lying in politics, though more attention has focused on the matter of the truth of peoples' beliefs and, in particular, the role of facts, evidence, impartiality, objectivity, emotions, identity, and reason in how they come to believe what they do. It is this latter sense of truthfulness, associated with what Bernard Williams (2002) called the virtue of 'accuracy', that we are interested in here. Issues such as how people come to their beliefs and their unwillingness to respond appropriately to evidence, traditionally of little interest to those outside of certain philosophical and psychological fields of study, have become deeply politicised as the battle between various parties has come to involve, often explicitly, competing claims as to who is the more truthful.

We can, broadly speaking, identify two sets of concerns regarding truthfulness. There are those anxieties often expressed under the unhelpful term 'post-truth', though discussed in more sophisticated

ways by Williams and Hannah Arendt before him (Arendt 1972; Williams 2002), that a sense of the value of truth is on the wane in Western liberal democratic societies (Ball 2017; D’Ancona 2017; Davies 2019; Haack 2019; Illing 2020; Kakutani 2018; Kavanagh and Rich 2018; Levitin 2017; MacMullen 2020; McIntyre 2018; Roberts 2020; Rosenfeld 2019; Stanley 2018; Zvi Baron 2018). And then there is a body of work that brings together research from the cognitive sciences, political psychology and democratic politics, that draws attention to the various biases and prejudices to which human cognition is prone and, in light of those, the difficulties in plausibly maintaining a vision of rational politics in which agents’ beliefs are determined by some ideal of impartial or objective reasoning (Achen and Bartels 2016; Bartels 2002; Bisgaard 2015; Galeotti 2018; Greene 2015; Haidt 2013; Kahan 2017; 2013; Kahneman 2012; Lodge and Taber 2013; Strickland, Taber, and Lodge 2011; Van Bavel and Pereira 2018). This second literature already provides us with plenty of grounds for thinking that politics is an area in which self-deception is rife, and, indirectly, why we should also therefore expect that many people are self-deceived about their truthfulness given that self-deception requires the agent be unaware of their epistemic failings.¹ But belief in our own truthfulness is a belief about ourselves and our own conduct analogous to, for instance, our belief in how courageous we are or how trustworthy, and hence it, and the possibility that it too is a matter on which we can be self-deceived, is something that we can explore more directly. Doing so is important for it allows us to focus on the ways in which how truthfulness features in our ethical self-image (individual and collective), and how it operates in political contexts, creates the conditions that make it a highly eligible candidate-belief for self-deception in its own right. If that is correct, then not only do we have further reason for thinking that self-deception in politics may be more common than we might like to think, it also helps us understand how it is that claims about truthfulness seem more likely to perpetuate and intensify conflicts in politics. Moreover, it suggests that the possibility of our being self-deceived about our truthfulness stands sufficiently independent of false first-order beliefs, and, as such, that it is likely to appear across the various political divides rather than being exclusive to one group. It is not

something that only afflicts ‘them’; we must be alive to the possibility that there is plenty of people on ‘our’ side who are self-deceived about their truthfulness also (potentially even including ourselves).

As a way of identifying the phenomenon of being self-deceived as to our own truthfulness (what we can call being ‘self-deceived about truthfulness’ for short) the first two sections of this paper cover the following: The first sets out the sort of scenarios or cases that are going to be candidates for self-deception about truthfulness, cases of ‘unwarranted belief in one’s truthfulness’ where the following is true of an agent:

- 1) They are genuinely committed to truthfulness;
- 2) They have gone about acquiring their beliefs (be they true or false) in ways that are inconsistent with what truthfulness requires;
- 3) They fail to recognise (2) and, as such, have an unwarranted belief that they have been truthful in their belief-acquisition process.ⁱⁱ

What then stands in need of explanation is how the agent can fail to recognise the shortcomings in how they have gone about acquiring particular beliefs given that they fall short of standards of truthfulness that they themselves hold.

The second section then sets out an account of self-deception that, in the third section, forms the basis of the argument that we can best make sense of cases of unwarranted belief in one’s truthfulness as instances where an agent is self-deceived as to how truthful they are. One aspect of this account will turn on the fact that we can be self-deceived about truthfulness only because it is something we value and, as such, we feel threatened, fearful or anxious when it is challenged. The third section will also expand upon this thought, providing reasons for thinking that there are individual and social dimensions of truthfulness as a value that enhance or amplify the likelihood that we might respond to such challenges through cognitive behaviours that lead to a self-deceived belief about our continued

truthfulness, and discuss the ethical and political consequences of this. The paper concludes with some wider reflections on truthfulness in the current political context.

Unwarranted Truthfulness

What is it to be committed to truthfulness? Where truthfulness is understood specifically in relation to the issue of accuracy we can say that it is to have the ‘desire for truth “for its own sake” – the passion for *getting it right ...*’.ⁱⁱⁱ Truth is directly connected to the notion of belief insofar as beliefs ‘aim at’ the truth, and hence accuracy is related to that aim of beliefs such that it ‘implies care, reliability, and so on in discovering and coming to believe the truth’ (Williams 2002, 126–27).

The direct connection between belief and truth draws our attention to two features of the virtue of accuracy. The first relates directly to the agent’s will, the second to the methods that agents use in their investigations. To take the second aspect first: some methods of inquiry are truth-acquiring in the sense that they have the genuine property, and which not all methods clearly have, of leading to true belief. Which method might be appropriate in this regard will depend upon the nature of the investigation itself, both what is being investigated and the knowledge one seeks. The general recognition that some methods track the truth better than others (while some do not do so at all) is part of what it is to appreciate how our beliefs are answerable to an order of things which lies beyond our determination and that may, in a variety of ways, resist being discovered. That is to say, there are external obstacles to acquiring true beliefs. The truth might be difficult to find or be hidden from us, require teamwork, very specific procedures, or advanced technologies to uncover, and part of what it is to exhibit the virtue of accuracy is to have the disposition to pick the appropriate method or approach for overcoming those obstacles. There are internal obstacles also to acquiring true beliefs, and the first aspect of accuracy relates to how any agent resists those.^{iv} These obstacles include the various ways in which an agent’s will – their attitudes, desires, and wishes - can hinder the acquisition of true beliefs.

The virtue of accuracy includes that array of skills, dispositions, and strategies which will sustain resistance against such internal obstacles to true beliefs as wishful thinking, laziness, accepting that which is agreeable or in some sense convenient to think true simply because it is agreeable or convenient, self-deception, and fantasy. This aspect of the virtue of accuracy is also directly related to appreciating how our beliefs are answerable to a reality that is 'out there' independent of us. Because beliefs aim at the truth, and what is true is determined by the nature of the way things actually are, one cannot bring it about that one believes p by willing that it be true. Indeed, were I to recognise the role that my will had played in the formation of my belief that p I would come to see that it has not been appropriately responsive to the world and hence that I was not warranted to hold it as a belief. 'My beliefs aim to be true, and, just for that reason, I must take them to be independent of my will' (Williams 2002, 135).

Someone who exhibits truthfulness has put their beliefs in the right relation to their will and the world. They should be reasonably satisfied that their beliefs are not only true but, as a condition of that, that they came to those beliefs in the right way, both in terms of the method they employed in their investigations and ensuring that they avoided the various ways in which their will can inappropriately impinge upon their belief-formation processes. It is, of course, perfectly possible that someone could be misguided in their sense of being satisfied that they have exhibited truthfulness. Indeed, the sort of cases we are interested in here are those in which an agent is genuinely committed to truthfulness and yet does not recognise that in acquiring certain beliefs they failed to meet the standards that they themselves believe truthfulness demands, a failure that leads them to be inappropriately satisfied that their beliefs are both true and that they came to them in the right way (or, we should say, true *because* they came to them in the right way). Maybe they picked an unsuitable method, they possibly succumbed to wishful thinking, or perhaps they failed to sufficiently check the findings of their investigations. Whatever the nature of their epistemic failings, they clearly undermine the grounds on which that person ought to consider themselves to exhibit truthfulness. If they continue to believe that they are truthful we are

entitled to say that they are no longer warranted in that belief – hence we can call these cases instances of ‘unwarranted belief in one’s truthfulness’.

The particular wrong involved in cases of unwarranted belief in one’s truthfulness relates to how the agent acquires their beliefs independent from whether the beliefs they come to hold are true or not. An immediate and unsettling social ramification of this is that even those who are on the right side of an argument might nevertheless still be guilty of holding an unwarranted belief in their truthfulness, with the epistemic failures that this implies. As a brief way of illustrating this possibility, let us introduce the ‘Remainer’.^v He is convinced that leaving the European Union will prove to be absolutely ruinous for the United Kingdom and that anyone who voted to leave the EU in the 2016 ‘Brexit’ referendum is guilty of having made a decision which will be a disastrous and wholly avoidable catastrophe for the country. The Remainer’s confidence in his political position is rooted not only in his sense that it is right but that he came to his belief in the right way. He sought out the relevant facts; he listened to the experts; he weighed up the evidence in an impartial manner; he followed the argument where it led him (and, he thinks, would lead any rational and good-willed person). Our Remainer would unlikely put it in these terms quite so explicitly, but he has an admirable commitment to truthfulness. Whatever the explanation for why Leave voters voted as they did, on any account that the Remainer finds plausible they failed to exhibit the virtue of accuracy.^{vi} They lacked the desire to know the facts as they really are and to ensure that their political commitments and beliefs were consistent with those. It is not simply that the leave voters were wrong, though he certainly thinks that, or that they were irrational, though that might have been the case also. Rather they did not care enough that their beliefs on which they acted were true. Our Remainer, by contrast, is satisfied not only that his beliefs are true but that he came to those beliefs in the right way, i.e. used the appropriate methods guided by the desire to ensure that they were consistent with or justified by the facts and best available evidence.

It turns out, however, that our Remainer is not quite as epistemically virtuous as he takes himself to be. The facts which he uses to justify his belief that there is overwhelming reason to support

the UK remaining in the EU he obtained through listening to the speeches and endorsements of pro-remain politicians, celebrities and businessmen, reading his friends' posts on social media (who tended to be as anti-Brexit as himself; he had blocked or un-friended any who were posting anything in favour of leaving), and reading his favourite (overwhelmingly pro-Remain) newspapers, blogs and websites. He did not seek to verify these sources; he took them at face value and never checked any facts or enquired as to whether his own media sources were reliable in the face of possible biases and agendas. He did not put any effort into searching for countervailing evidence or arguments or to ever check whether these 'facts' were indeed correct. In coming to his decision how he would vote in the referendum he did not, and has not since, subjected his beliefs to any level of critical scrutiny. Not only did he not seek to verify the facts that he took to be true, neither did he worry whether his commitment to EU membership might in any way be influenced by his highly pro-European sense of self-identification, his vehement dislike of the politicians who campaigned for Brexit, or his strong desire to be able to move freely around Europe on his regular lengthy luxurious holidays on the continent.

Let us grant, for the sake of argument, that the facts are indeed as the Remainer takes them to be, and that they do conclusively demonstrate that he is right in his opposition to Brexit such that any rational and impartial agent should have seen the overwhelming merits of the case for remaining within the European Union. Even with that granted, it is important that if the Remainer is right about Brexit it is despite and not because of the ways in which he has gone about acquiring his beliefs. It is not because he has done a better job of having ascertained the facts than the Leavers he criticises. Hence our Remainer is culpable of a significant failure of truthfulness, and, importantly, that is the case even if his beliefs turn out to be true. In that sense, his (unwarranted) belief in his truthfulness stands independent from the truth or falsity of his first-order beliefs about Brexit.

How are we to make sense of the Remainer's unwarranted belief in his truthfulness? One thing we might want to say is that the epistemic failures in how he went about acquiring his beliefs about Brexit gives us good reason to doubt that he really is as admirably committed to truthfulness as we have

allowed. After all, we might think that it is much more plausible to infer from the fact that someone does not go about acquiring their beliefs truthfully that they therefore lack the commitment to truthfulness, despite their belief and any protestations that they do.^{vii} In some cases this might well be the right inference to make. Alternatively, we might want to say something like the Remainer does not properly understand what truthfulness actually requires, questioning not so much his commitment to truthfulness but his notion of what it demands in terms of how we go about acquiring our beliefs. But we have said of the Remainer that his commitment to truthfulness is both genuine (he really is committed to truthfulness) and that he has a good grasp of what it means, so neither of these explanations will suffice. The question then remains how an agent can plausibly continue to believe that they are truthful when the ways in which they have acquired their beliefs fall short of what they themselves (rightly) believe truthfulness demands? The best explanation for this is that the agent's continued belief in their truthfulness is a form of self-deceit.

Self-Deception

One familiar way of thinking about self-deception, and one common to a particular strand of the philosophical literature on the topic, is to conceive of it as a first-person form of the third-person phenomenon of deception. To deceive, on this account, is to intentionally bring another to believe something that you know is false; to be *self*-deceived is to intentionally bring *yourself* to believe something that you know is false.^{viii} That this way of understanding self-deception generates numerous philosophical paradoxes or difficulties is hardly surprising, and much of the literature is dedicated to the attempt to resolve them.^{ix} And yet this way of thinking about self-deception is misleading. Just as it would be wrong to think that it is paradigmatic of deception that it be intentional (one can unintentionally deceive others), we should resist the assumption that all self-deception must be intentional also (Mele 1997; Lazar 1999; Oksenberg-Rorty 1985; Mele 1987).^x Moreover, it is not

right that self-deception (nor third-person deception) requires one to change existing beliefs. Few cases of self-deception follow the structure of bringing oneself to believe $\sim p$ when we already believe p to be true (or vice versa), not least because much self-deception relates to matters on which we had no pre-existing beliefs in the first place (Mele 1997; 2001). The parent need not have initially believed that their child does take drugs in order to be self-deceived in their belief that they do not. At best, to think of self-deception as a first-person form of third person deception results in us mistaking a part of the phenomena for the whole; at worst, it misleads us as to the nature of the phenomena itself.

Though usually unintentional, self-deception is not accidental.^{xi} I can accidentally cause myself to hold false beliefs via a myriad of different routes, such as if I am careless when reading a document or misrecognise a particular landmark when trying to find my way somewhere, but I am not self-deceived in holding the resulting beliefs. Likewise, we know that human cognition is liable to a whole host of behaviours that can lead people to reach biased and false conclusions, but we tend not to think that they are self-deceived in holding those beliefs. Or, at least, we do not where those beliefs are unmotivated or ‘cold’ in the sense that the cognitive behaviours have acted independently from or without their being triggered by some causal motivation. Motivated reasoning, on the other hand, occurs where one holds a belief p that has been causally sustained or produced by a motive for believing that p be true.^{xii} One prominent way of thinking about self-deception is as a species of motivated reasoning in which an agent is motivated (by their desires, wishes, fears, interests, emotions) to produce or retain a false belief where that belief is held in the face of weightier evidence to the contrary.^{xiii} Beliefs sustained in such a manner are ones that we are self-deceived in holding. Yet the role that the underlying motivation has played in leading the agent to the false belief will not be transparent or conscious to them. They need not set out with the explicit intention that they will bring about that they believe p , though neither is it merely a matter of chance or misfortune that they have done so – their coming to believe p is in part a function, even if they do not realise it themselves, of their wanting to believe p .

The desire for p to be true – because I wish it to be so or because it might somehow serve my interests were it so – can provide the motivation that produces or sustains the belief that p . There are a variety of ways in which our desires and interests can contribute to our belief formation-processes in general but also in cases of self-deception. Alfred R. Mele lists four possibilities, all of which he takes as being involved in ‘garden-variety’ forms of self-deception (Mele 1997, 94; 2001):

Negative misinterpretation: Our desiring that p may lead us to misinterpret as not counting (or not counting strongly) against p data that we would easily recognise to count (or count strongly) against p in the desire’s absence.

Positive misinterpretation: Our desiring that p may lead us to interpret as supporting p data that we would easily recognise to count against p in the desire’s absence.

Selective focusing/attending: Our desiring that p may lead us both to fail to focus attention on evidence that counts against p and to focus instead on evidence suggestive of p .

Selective evidence-gathering: Our desiring that p may lead us both to overlook easily obtained evidence for $\sim p$ and to find evidence for p that is much less accessible.

In each of these cases the desire that p be true motivates defective reasoning which results in the false belief p . To give two examples of how this can lead specifically to self-deception: (1) A man is fired from his job on the basis of misconduct and gross negligence with immediate and devastating financial implications for his family. When faced with the clear evidence of his transgressions he refuses to accept that they amount to misconduct and grounds for dismissal, even though he himself had fired an employee for the same actions in a previous job (negative misinterpretation); (2) A political scientist has made a significant name for themselves in the profession as a staunch advocate of a particular theory that

x will not happen in conditions y . When x does happen in condition y , the political scientist works very hard to maintain, on slender and dubious evidence, that y never actually pertained in that case and hence that their theory has not been debunked, overlooking the clear and readily available evidence to the contrary (selective evidence-gathering). These are both cases in which we can say that the agent is motivated to believe p or $\sim p$ by the fact that that is the way they *want* the world to be. Again, this need not be interpreted intentionally, but it does mean that we can say that their belief has tracked the way they wish reality was rather than the way that it happens to be.

While we know that the human capacity for self-deception is severely limited (Kunda 1990), there are likely still a great number of false things which we might wish or desire to be true and yet we not only do not bring ourselves to believe but *could not* bring ourselves to believe. That is to say, not all false beliefs seem to be candidates for self-deception. There is a selectivity to what it is we can be self-deceived about. At least part, and potentially a large part, of the story regarding how that selectivity actually functions will likely focus on the role of the emotions. It is telling that it is hard to think of a typical case of self-deception in which the false belief is not associated with a strong emotional response (Dagleish 1997; Lazar 1999; Lodge and Taber 2005; Galeotti 2018).^{xiv} Indeed, the stock example of self-deception in which a spouse is unwilling to believe that their partner is cheating on them in the face of strong evidence rings true to us as an instance of self-deception precisely because we recognise the high-level emotional stakes involved in facing up to the truth (the pain of accepting that the person you love has betrayed you, the humiliation of the affair being public knowledge, the fear that you won't be able to manage on your own, the consequences for the wider family, etc.). Galeotti is on the right lines when she points out that typical examples of self-deception tend to do with what she calls “mortal questions”, that is matters which bear a fundamental and constitutive relationship with the self ...'. These include marital infidelity, fatal illness, and a child's problems with addiction or criminal behaviour where the truth is likely to arouse feelings of great fear and anxiety. In such cases, the wish or desire that p is threatened by negative evidence and is so vital in fulfilling a crucial desire-element for an

agent's well-being and prospects that it becomes 'emotionally overloaded', and in doing so triggers the cognitive behaviours that then lead to self-deception (Galeotti 2018, 44–45). This offers a plausible explanation of how it is that self-deception can operate selectively as opposed to in relation to all of our desires or interests. So while we can readily understand how beliefs regarding the non-infidelity of one's spouse can become 'emotionally overloaded' and hence the subject of self-deception, in contrast, we find it hard to imagine a scenario in which someone is self-deceived about which bank offers the best customer service, the finest train route, or the most durable grass, because it is difficult to envisage the answers to these questions ever being of such significance that they arouse the sort of emotions (of desire, fear, or anxiety) that could motivate the processes involved in self-deception.^{xv}

Evidence that is interpreted as threatening notions of identity and well-being can often also generate substantial emotions of fear and anxiety and hence are likely candidates for self-deception. We know that when it comes to evidence that threatens the self, in particular whatever we take to be integral features of our self-image, humans have the capacity to re-describe unwelcome truths in ways that allows us to maintain that self-image. Galeotti (2018, 39) cites experiments by Wentura and Greve (2003; 2005) in which participants who thought of themselves as cultured by virtue of their historical knowledge fail a history test in the context of the experiment and yet find ways to process that negative result by adapting what they thought was required to qualify as cultured, either by questioning the value of the test or by denying their (previously stated belief) that knowledge of history was a condition for being cultured. If notions of being cultured are sufficient to motivate self-deception then it is not hard to see how protecting other aspects of one's identity could potentially activate the processes of self-deception also, such as when one might be unable to recognise the wrongs perpetrated by one's country in a conflict, their family's role in historical injustices, or the good policy decisions being taken by a despised governing party.^{xvi}

We finally need to make the significance of evidence to self-deception explicit. It is not sufficient for something to count as an instance of self-deception that the relevant belief has been

produced or sustained through motivated reasoning. That, after all, is true also of wishful thinking. In cases of wishful thinking, the route from desire to belief is, so to speak, direct. The desire causes the belief, independent of any evidence for against the desired conclusion. To be guilty of wishful thinking is to believe p because I desire p to be the case. And it is precisely in having one's will determine what you believe to be true rather than consideration of the way the world really is that the problem of wishful thinking lies. Self-deception is where desires and emotions have improperly affected the way in which the agent assesses evidence for a belief. The partner who cannot bring themselves to believe their spouse's infidelity fails to follow the evidence to the proper conclusion; the jealous lover over- or misinterprets certain actions or events as evidence of their partner's exploits. Likewise, each of Mele's 'garden-variety' forms of self-deception are instances where an individual's desire has led them to either misinterpret the evidence or to go astray in how they attend or collect evidence in ways that would not be the case if the desire were not present. The route from desire or emotion to belief is indirect in cases of self-deception in that they cause the agent to hold the false belief by virtue of hindering or impeding their assessment of evidence in rational belief formation. Proper attributions of self-deception are going to turn, therefore, and in a way not true of cases of wishful thinking, on judgements about the inappropriate causal role that desires and emotions have played in how an agent has assessed evidence in coming to hold the false belief.

In imputing self-deception we are making a judgement not only about the improper ways in which an agent has come to hold a belief but regarding the nature of the evidence available to the agent also. Someone cannot plausibly be self-deceived if the evidence is genuinely inconclusive such that people can reasonably interpret it as supporting different beliefs. The same is true if the evidence is in some important sense misleading or incomplete such that the agent can reasonably be misled in the beliefs that it does and does not support. We cannot plausibly, for instance, criticise an agent for holding self-deceived belief p when they did not have access to the evidence and information that would lead them to believe $\sim p$, or if the partial evidence that they could access pointed more in the direction

of p over $\sim p$. Rather, it must be the case that the evidence is overwhelming such that it is reasonable to expect any rational agent to infer p over $\sim p$ from it.^{xvii} To be self-deceived is to not only believe p because of the desire or wish for p to be true but to do so when p is not the belief most strongly supported by the relevant available evidence.^{xviii}

Self-Deception about Truthfulness as an Ethical (and Political) Problem

Can we make sense of cases of unwarranted belief in one's truthfulness in terms of self-deception?

There is a very straightforward but important sense in which self-deception and wishful thinking about first-order beliefs are inevitably going to produce second-order self-deception about truthfulness as a by-product. The indirect process that generates this can be reconstructed as follows: a) A badly want that p , despite the available evidence; b) A unintentionally sets an inaccurate cognitive process in motion which results in the belief that p ; c) A is candidly convinced that p is true; d) As a by-product, A is also convinced that they have come to p truthfully, and; e) As a result, A is self-deceived about p , and self-deceived about being truthful about p . The inaccuracy of the epistemic processing in cases of self-deception will always be opaque to the agent, ensuring that a continued and unwarranted belief in one's truthfulness must inevitably accompany instances where we are self-deceived about our first-order beliefs. Were the epistemic failures transparent to the agent then they would no longer be able to hold the self-deceived belief that p . Wherever there is wishful thinking or self-deception we will find self-deception about truthfulness also.

This account explains unwarranted belief in one's truthfulness in those cases where false first-order beliefs have been arrived at via wishful thinking or self-deception. As such, it will help us understand those cases where the agent's first-order beliefs are false. But we allowed, in the case of the Remainer, that his first-order beliefs about Brexit were correct. The reason why we can still say that he had an unwarranted belief in his truthfulness is due to the faulty way he came to acquire those beliefs

which happened, fortuitously, to be true. That rules out being able to explain his unwarranted belief in his truthfulness as an indirect by-product of being self-deceived as to the truth of his first-order beliefs. A different explanation is therefore required. We also ought to allow for the possibility that our beliefs about our truthfulness are not just simply determined by how we come to acquire our first order beliefs, as if there are not other considerations which press upon the question of how truthful we are more directly and which might have some role to play in leading us to be self-deceived as to our truthfulness (some of which we will go on to explore in this section). If that is possible then the causal arrow might point in the opposite direction: because we are self-deceived as to how truthful we are we therefore overlook our epistemic failings in acquiring our first order beliefs, be they true *or* false. So, do we have reason to think that we can be self-deceived as to our truthfulness for reasons independent of how we came to our first-order beliefs? We do if it is possible to plausibly describe the continued belief in the agent's truthfulness as the product of motivated belief formation and can identify credible causal motivations that could have triggered the relevant cognitive behaviours. Let us start with the former condition first.

Recall two of the four mechanisms referenced earlier as involved in many ordinary and familiar cases of self-deception - selective focusing/attending and negative misinterpretation – and let us consider how the unwarranted truthfulness of our Remainer might be explicable as the result of these:

Selective focusing/attending: Because of his desire to believe that he is truthful, the Remainer failed to focus attention on the evidence that counts against his truthfulness, such as that he consulted sources from too limited a range of perspectives, or that he did not put his beliefs up to any level of critical scrutiny. He also did not consider the extent to which his existing desires and wishes, as well as he strong pro-European identity, might have created a strong predilection for him to be more naturally inclined to pro-Remain messages which he ought to have compensated against by verifying or fact-checking the information he received.

Instead, our Remainer focused on evidence that he took as suggestive of his truthfulness, such as the fact that he has consulted a large number of sources, or that his own experiences of travelling around Europe gave him an insight into the value of EU membership for the UK.

Negative misinterpretation: Because of the Remainer's desire to believe that he is truthful he misinterpreted the various inadequacies of his belief-acquisition processes as not counting (or not counting strongly) against his truthfulness in ways that he would easily recognise to count (or count strongly) against his truthfulness in that desire's absence, or which he would point to if assessing the failing of others' (for example, Leavers') belief-acquisition processes. That he would interpret the inadequacies in this way is evident from the fact that he takes it to be a serious failing of Leave voters that their beliefs were (he believes) arrived at through deeply flawed processes or derisory investigation which undermine their truthfulness.

In the case of selective focusing/attending the Remainer fails to focus his attention on the relevant evidence. In that of negative misinterpretation the Remainer does attend to the evidence but misinterprets it as not counting against his truthfulness. Via either or both routes he is led to the false belief that he is truthful by his desire to believe that he is truthful, contrary to the available evidence.

In that sense, the cognitive behaviours that could lead one to be self-deceived about truthfulness are not necessarily hard to fathom. Indeed, beliefs about our truthfulness are beliefs like any other, and as such we need not be looking for special or idiosyncratic explanations as to the mechanisms that could lead to a false rather than true conclusions. Where more is needed to be said is in relation to the issue of selectivity, whether it is plausible that the desire to believe in our own truthfulness can be the sort of belief that triggers those cognitive behaviours in the first place. And here we have reason to believe that truthfulness may almost be overdetermined as a candidate-belief for self-deception.

It may not be universally true, but it matters to many of us that we are able to see ourselves as truthful, that we stand to our beliefs as true beliefs with all that that entails. More than that, it matters that it be true that we *are* truthful. We feel the pull of truthfulness and recognise the significant demands that it places upon us, not least when we come to see that the pursuit of truth may require us to believe that which we really would rather not. Even if one believes they have never succumbed to the temptation to believe the convenient over the true, however credibly, the very recognition of the considerable tension and the cognitive and moral dissonance that it threatens signals the significance that we give to the importance of holding true beliefs. And this is not just for instrumental reasons. Most of us would shirk at the thought that in forming our beliefs we did so, even if unintentionally, in a way that was guided away from the true by our desires or wishes. We would, in an important sense, take it to be an assault on our self-image; a failing of character along the lines of a failure of courage, which, in an important sense, it is.

After all, we very often (and revealingly) take it to be a character failure when we think we see it in others. This is sometimes explicable through a simple distinction: we know there is a difference between a truth too difficult to accept and a truth too inconvenient to accept. Accepting truths of the first kind often requires a strength of courage that few of us are regularly called upon to muster, and we can readily understand how it might prove beyond us in the most terrible of circumstances ('I just cannot believe my partner has died'). Nonetheless, we must be careful not to undervalue the courage needed to even accept truths of the second kind, at least in some circumstances. We may often be too hasty in scorning or ridiculing those who we think are guilty of believing the convenient over the true by, for instance, accusing them of failing to recognise that they are not the centre of the universe, or that the world is not designed to fulfil their purposes, and which we associate with a mature or disenchanted stance towards reality. In truth, it is often a lot more difficult to be truthful than we are apt to credit others with in our judgements of them – which, of course, is *itself* a form of self-deception, and untruthfulness, in turn. As we have seen, the obstacles to achieving such a stance can be great

indeed, and we must not pretend that the realisation that our beliefs ought to be responsive to the world and not our will is always an easy part of our psychological development (or one that everyone completes), or that it alone guarantees our being truthful. Precisely because of this we are therefore on ethical grounds when we are involved in the attempt to ensure that our beliefs stand in the right relationship to reality. The question of our truthfulness becomes part of our identity and, crucially, our own sense of our ethical standing, as well as of the ethical judgements we are passing on others and ourselves. Where this is the case, we place a high value on our sense of ourselves as living up to the demands that truthfulness places upon us. Where we fall short we will recognise this as a profound failing; where we meet those demands we will recognise that as the ethical achievement it is.

It is easy to see from this that to have our truthfulness queried, by ourselves or others, is liable to be interpreted or experienced as threatening an integral feature of our ethical self-image as truthful – one that therefore generates emotions of fear and anxiety which can trigger the cognitive behaviours that enable us to retain our belief in our truthfulness (in the face of the evidence that we are not). Where we place a high value on truthfulness we will possess the underlying latent motivation (to be truthful) that would make it possible to deceive ourselves as to how far we truly embody its virtues. And so, to return to the case of our Remainer, by deceiving himself as to the extent to which he has acquired his beliefs about Brexit in the right way, as truthfulness requires, and how far he has engaged in the sort of critical reflection upon his beliefs that he believes truthfulness demands, *he is able to retain his sense of himself as truthful, as someone who seeks to relate to himself, others, and the world, in a truthful manner.* He is self-deceived about his truthfulness, and can only be self-deceived in this way, precisely because it occupies such an important space in his ethical self-image. Yet the political upshot of this – noting that the exact same process is likely to be happening with Leavers on the opposite side of the ideological fence – is that challenges issued from one side to the other over matters of truth are necessarily interpreted not as sincere invitations to consider relevant facts, even when they are intended as such, but as *attacks on one's fundamental ethical integrity by their opponents.* Such attacks are therefore

experienced as a provocation, and we should be entirely unsurprised that in highly charged political disagreements, appeals to be truthful enflame, rather than, dampen political conflict.

Yet challenging one's truthfulness not only directly confronts a particular aspect of our ethical self-image, it has the potential to destabilise our first order beliefs also insofar as the potential that we are not as truthful as we have believed ourselves to be would give us reason to suspect that in acquiring our other beliefs we might not have always successfully countered the various obstacles to discovering the truth there either. A pervasive self-doubt that permeates through all of our beliefs may loom. In addition to the concern that this exposes our failures of truthfulness, it also raises the possibility that we are no longer warranted in holding in particular those cherished first order beliefs that are central to our self-identity. For our Remainer for instance, someone who considers their political and social beliefs to be crucial facets of their identity, these would include, among other things, their beliefs about Brexit. To countenance the possibility that he is not truthful is to countenance the possibility that the beliefs he holds as in some crucial sense constitutive of who he is might be false, and he might in turn have to change who he feels he most fundamentally *is*. Few of us are comfortable with the prospect of that, and understandably so. What these beliefs are which we take as having a constitutive place in our self-identity will, of course, vary between individuals, but regardless of the content of the belief it is easy to see how a challenge specifically to our truthfulness can, in addition, be interpreted and experienced as a threat to those beliefs also, beliefs we find it incredibly difficult, maybe even painful, to doubt. Continuing to falsely believe in our truthfulness gives us one way of shielding those beliefs from such suspicion.^{xix} The effect of this, one would imagine, is to significantly heighten both the likelihood and extent to which challenges to our truthfulness will be interpreted or experienced as threats to the self that generate the sort of emotional response which makes self-deception a possible effective response.

The account of self-deception that we have relied upon here has had a strong individualist bias, and several authors have rightly urged the need to take more seriously the social dimensions of self-deception (Harré 1988; Ruddick 1988; Solomon 2009), in particular in helping maintain and support

individual instances of self-deception. And thinking about self-deception as a social phenomenon allows us to supplement what we have already said about the propensity to be self-deceived as to our truthfulness.

Where certain beliefs are considered to be an essential element or qualification of group membership, either in the sense that the group is defined by what it believes or where holding certain beliefs is necessary for qualification into that group, then there is a collective interest in mutually reinforcing each other's confidence in their truthfulness so as to stave off the possibility of doubt concerning the veracity of those constitutive beliefs creeping in. Members of such groups have a mutual interest in collectively supporting their individual sense of truthfulness. There are a variety of dynamics that can work to this effect. One of these includes developing strong local notions of epistemic authority or expertise and the corollary grounds of that authority, which will include, if only implicitly, an account of why it is that their group has particular or unique access to the truth (e.g. 'we follow Trump because he is an outsider and so can see things for what they really are'; 'those on the right are unable to see the truth because they are unwitting subjects of the neoliberal hegemony which we on the left recognise and reject'). A different and often very effective dynamic includes disparaging the dissenting beliefs of others and other groups by questioning not the content of what they believe but their very commitment to truthfulness in the sense of caring that their beliefs are true and undertaking the sort of due diligence to ensure that they are. Our Remainer, for instance, cannot contemplate the possibility that Leave voters may have a legitimate position but is only able to understand their beliefs as the result of profound failings of truthfulness, and that functions as a way of him both identifying strongly with those with likeminded beliefs and bolstering his/their own sense of truthfulness. Groups can therefore feel threatened by challenges to their truthfulness insofar as it calls into question the basis of their collective self-image. And where that is the case, it is not only plausible to imagine that this creates a sense of fear and anxiety that can trigger self-deception about truthfulness as a response, but that there are a host of self-enforcing mechanisms that groups have at their disposal to uphold and support that

individual and collective self-deceit. Once again, therefore, we should not be surprised to find that demands to be truthful have recently proved to be politically incendiary, not pacifying.

There is a much broader social dimension here that needs to be factored in also. It is central to a prevalent and pervasive self-image of modern Western civilisation that it gives great weight to the value of truthfulness. It is, on this narrative, part of what defines us, both historically speaking from our own past and from other civilisations. This is a vastly complex issue, but two very brief and related remarks might suffice to establish the point. Unintentionally or not, the thinkers of the Enlightenment era ushered in a new secular age, one that abandoned a theocentric understanding of the universe and our place within it and replaced it with various alternatives that put great emphasis on the powers of reason and science to discover the vital truths of humanity and nature emancipated from superstition and false reason.^{xx} These would then form the basis of new frameworks of knowledge and visions of human society. In its crudest, which is not to say uninfluential, form this narrative is a straightforward story of progress, literally of enlightenment, in which this historical development represents an advancement from a time of superstition and ignorance to one of light and understanding, and advancement of truthfulness. A different though not unrelated narrative, represented in varying ways by thinkers such as Marx, Nietzsche, Freud and those associated with the Frankfurt School, is deeply suspicious of the claims made on behalf of reason and science in modernity precisely on the grounds that they have not pushed the will to truthfulness far enough. These misgivings can, when pushed too their extreme, lead to severe scepticism about the very existence of truth or our possibility of knowing it, though more often they serve as motivations to seek and identify deeper truths which have up until now (allegedly) been obscured by power. This narrative is often thematically associated with disenchantment or unmasking, the chief idea being that we have, finally, thrown off even the reassuring myths and comforting falsehoods that post-Enlightenment modernity relied upon. It is a disenchanting that can often be deeply disconcerting and almost always challenging (which is why Kant exhorts us not just to know but to *dare* to know), though it is one undertaken in the name of truthfulness (something

Nietzsche, if not necessarily his followers, was acutely aware of). There is, of course, a question about the truthfulness of these narrative themselves, and it is a long-standing question the extent to which they have been sustained only by systematically oppressing and excluding other valid forms of knowledge that are incompatible with it. But in that sense even its critics recognise and confirm the centrality of truthfulness to the Western self-image.

Insofar as such narratives which valorise truthfulness are part of the self-image of modern Western civilisation, they embed themselves into our own sense of individual identity also. It is vital to one's own self-identity that *I* am truthful because it is vital to our collective identity that *we* are truthful. That is a fundamental part of what it means to be a member of Western civilisation itself, at this point in our historical development. A threat to my truthfulness is thus not experienced as an isolated challenge to some idiosyncratic individual commitment, but rather speaks to a wider collective ethical self-image in which we understand ourselves and hence raises questions as to our standing in relation to it. This further compounds the potential sense of fear and anxiety that challenges to our truthfulness might create and makes it more likely that an agent responds by continuing in the false belief in their truthfulness through self-deception. It also raises again the likelihood of a social response to such challenges that does not seek to support truthfulness through its practices and institutions, but rather works to aid each other in our own ongoing self-deceptions. Rather than disrupt others' self-deceptions regarding their truthfulness, we might encourage it precisely so we can perpetuate our individual and collective self-deceits. Indeed, because it matters that others recognise our truthfulness given its social dimensions, and we know that others who value truthfulness seek such recognition also, mutual self-deception might be both easy and attractive. Self-deception is, after all, simpler when others are able to help deflect scruples, avoid inconvenient evidence, or reinforce the right reassuring messages. When our political opponents accuse us of betraying the virtue of truthfulness, the default will not be to interpret this as an invitation to dispassionate rational enquiry, but as a profound ethical – and, in relevant contexts, therefore also a political – provocation.

Conclusion

The possibility that we have been exploring here is that it is both possible that we can be self-deceived about our own truthfulness and that there is something about the value of truthfulness which makes it a strong candidate for such self-deception, and especially so at this historical point in Western culture. As we have seen, the individual and social dimensions of truthfulness amplify the potential for the sort of emotional response necessary to prime the cognitive behaviours that can lead to self-deceived beliefs.

There are at least two reasons why this should be a particularly disquieting possibility. The first is that it lends further support to the possibility there simply may be a lot of it going on. Moreover, and this feels particularly pertinent today, self-deception about truthfulness is not the same as self-deception about particular political beliefs (though as we have seen there is clearly a relationship between them), and hence we should expect that being self-deceived as to our truthfulness is a phenomena that straddles all political divides, whilst also making such divides grow wider. Indeed, because one can be self-deceived about truthfulness even if one's other beliefs happen fortuitously to be true, then we have to be open to the possibility not only that those who share our beliefs might not be as truthful as they think or present themselves. It might, and this is the most unsettling thought of all, jolt us out of any personal complacency that *we ourselves* are as truthful as we like to believe we are or that we have always come to our beliefs in ways consistent with the demands of truthfulness. Insofar as politics has in recent years become deeply destabilised over, in part, claims to truthfulness and accusations of the opposite, one implication of this paper is that the problem can never just be 'them', it also has to be 'us', too – and that is so even if we happen to be right.

Second of all, the possibility being considered here complicates the challenge of defending truthfulness considerably. The main strategy pursued by defenders of truth in their various guises today has been to remind people of its value. That is only a feasible solution if the problem is that people have

or are at risk of forgetting its value. If the problem, rather, is that people remain committed to the value of truthfulness but can be self-deceived as to how far that they are themselves acting truthfully, then the response must clearly go beyond reiterating to those people the demands of truthfulness that they already sufficiently understand. The task, rather, must be to continue to buttress and defend those institutions and practices that support truthfulness in liberal democratic societies, whatever those might be, but in such a way that does not at the same time encourage us to self-deception given, as we have seen, our propensity to be self-deceived about truthfulness is enhanced by the special place it has in our ethical lives, individually and collectively. Part of what we recognise in the self-deception of the cuckolded partner is that it is their love which has gone awry in leading them to abandon care for the truth of their beliefs. It is genuine love they have for their partner, for if it were not then it is unlikely that it could motivate the self-deception in the first place. But we tend to think that it is in some sense a distorted love, one that has become dangerous through its intensity in leading the agent to false beliefs by untethering them from any concern for truthfulness. And maybe the same is true in the other direction also, that a truthfulness which is not understood in connection with, and as in various ways contained within, its healthy limits by other values is likely to be something over which we will be prone to be self-deceived. The trouble is that in an intellectual tradition that has vacillated between thinking that truth ought to be the ultimate guide in politics, or provide the philosophical foundation upon which stable, rational, or true political orders can be built, versus thinking that truth has no place in politics or that it is essentially impotent in the chaos and conflict of political life, has meant that we still do not have a very clear understanding of what those connections are, or what we need them to do for us if we are to live well together.

Bibliography

Achen, Christopher H., and Larry Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Oxfordshire: Princeton University Press.

- Arendt, Hannah. 1972. *Crises of the Republic*. London: Harcourt Brace.
- Baghrarian, Maria, and Anna Nicholson. 2013. 'The Puzzle of Self-Deception'. *Philosophy Compass* 8 (11): 1018–29.
- Ball, James. 2017. *Post-Truth - How Bullshit Conquered the World*. Great Britain: Biteback.
- Bartels, Larry M. 2002. 'Beyond the Running Tally: Partisan Bias in Political Perceptions'. *Political Behavior* 24 (2): 117–50.
- Bisgaard, Martin. 2015. 'Bias Will Find a Way: Economic Perceptions, Attributions of Blame, and Partisan-Motivated Reasoning during Crisis'. *Journal of Politics* 77 (3): 849–60.
- Dagleish, Tim. 1997. 'Once More with Feeling: The Role of Emotion in Self-Deception'. *Behavioral and Brain Sciences* 20: 110–11.
- D'Ancona, Matthew. 2017. *Post-Truth - The New War on Truth and How to Fight Back*. London: Penguin.
- Davidson, Donald. 1986. 'Deception and Division'. In *Problems of Rationality*, 199–212. Oxford: Clarendon Press.
- Davies, William. 2019. *Nervous States: How Feelings Took Over the World*. London: Vintage.
- Demos, Raphael. 1960. 'Lying to Oneself'. *Journal of Philosophy* 57 (18): 588–95.
- Dolan, R. J. 2002. 'Emotion, Cognition, and Behavior'. *Science* 298 (5596): 1191–94.
- Druckman, James N., and Toby Bolsen. 2011. 'Framing, Motivated Reasoning, and Opinions about Emergent Technologies'. *Journal of Communication* 61 (4): 659–88.
- Druckman, James N., James H. Kuklinski, and Lee Sigelman. 2009. 'The Unmet Potential of Interdisciplinary Research: Political Psychological Approaches to Voting and Public Opinion'. *Political Behavior* 31: 485–510.
- Enns, Peter K., and Gregory E. McAvoy. 2012. 'The Role of Partisanship in Aggregate Opinion'. *Political Behavior* 34: 627–51.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. CA: Stanford University Press.
- Galeotti, Anna Elisabetta. 2018. *Political Self-Deception*. Cambridge: Cambridge University Press.
- Greene, Joshua. 2015. *Moral Tribes - Emotion, Reason, and the Gap between Us and Them*. Great Britain: Atlantic Books.
- Haack, Susan. 2019. 'Post "Post-Truth": Are We There Yet?' *Theoria* 85: 258–75.
- Haidt, Jonathan. 2013. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. London: Penuin Books.
- Harré, Rom. 1988. 'The Social Context of Self-Deception'. In *Perspectives on Self-Deception*, edited by Brian P. McLaughlin and Amelie Oksenberg-Rorty, 364–79. London: University of California Press.
- Illing, Sean. 2020. "'Flood the Zone with Shit": How Misinformation Overwhelmed Our Democracy'. Vox. 6 February 2020. <https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation>.
- Izard, Carroll E., Jerome Kagan, and Robert B. Zajonc. 1984. *Emotions, Cognition, and Behavior*. Cambridge: Cambridge University Press.
- Kahan, Dan M. 2013. 'Ideology, Motivated Reasoning, and Cognition Reflection: An Experimental Study'. *Judgment and Decision Making* 8: 407–24.
- . 2017. 'Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition'. *Cultural Cognition Project Working Paper Series*, no. 164. <https://ssrn.com/abstract=2973067>.
- Kahneman, Daniel. 2012. *Thinking, Fast and Slow*. London: Penguin Books.
- Kakutani, Michiko. 2018. *The Death of Truth*. Great Britain: William Collins.
- Kavanagh, Jennifer, and Michael D. Rich. 2018. 'Truth Decay - An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life'. RAND Corporation.
- Kruglanski, Arie W. 1989. 'The Psychology of Being "Right": The Problem of Accuracy in Social Perception and Cognition'. *Psychological Bulletin* 106 (3): 395–409.
- Kunda, Ziva. 1990. 'The Case for Motivated Reasoning'. *Psychological Bulletin* 108 (3): 480–98.

- Lazar, Ariela. 1999. 'Deceiving Oneself or Self-Deceived? On the Formation of Beliefs "Under the Influence"'. *Mind* 108: 265–90.
- Levitin, Daniel. 2017. *Weaponized Lies: How to Think Critically in the Post-Truth Era*. Great Britain: Penguin Books.
- Lodge, Milton, and Charles S. Taber. 2005. 'The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test'. *Political Psychology* 26 (3): 455–82.
- . 2013. *The Rationalising Voter*. Cambridge: Cambridge University Press.
- Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. 'Biased Assimilation and Attitude Polarisation: The Effects of Prior Truth Theories on Subsequently Considered Evidence'. *Journal of Personality and Social Psychology* 37 (11): 2098–2109.
- MacMullen, Ian. 2020. 'What Is "Post-Factual" Politics?' *Journal of Political Philosophy* 28 (1): 97–116.
- McCabe, Katherine T. 2016. 'Attitude Responsiveness and Partisan Bias: Direct Experience with the Affordable Care Act'. *Political Behavior* 38: 861–82.
- McIntyre, Lee. 2018. *Post-Truth*. USA: MIT Press.
- Mele, Alfred. 2003. 'Emotion and Desire in Self-Deception'. *Royal Institute of Philosophy Supplement* 52: 163–79.
- Mele, Alfred, M. 1987. *Irrationality - An Essay on Akrasia, Self-Deception and Self-Control*. Oxford: Oxford University Press.
- . 1997. 'Real Self Deception'. *Behavioral and Brain Sciences* 20: 91–136.
- . 2001. *Self-Deception Unmasked*. Princeton: Princeton University Press.
- . 2004. 'Motivated Irrationality'. In *The Oxford Handbook of Rationality*, 240–56. Oxford: Oxford University Press.
- Niedenthal, Paula M., and Shinobu Kitayama. 1994. *The Heart's Eye: Emotional Influences in Perception and Attention*. San Diego: Academic Press.
- Oksenberg-Rorty, Amelie. 1985. 'The Deceptive Self: Liars and Layers'. *Analyse & Kritik* 7: 141–61.
- Pears, David. 1984. *Motivated Irrationality*. Oxford: Oxford University Press.
- Pietryka, Matthew T., and Amber E. Boydstun. 2012. 'Going Maverick: How Candidates Can Use Agenda-Setting to Influence Citizen Motivations and Offset Unpopular Issue Positions'. *Political Behavior* 34: 737–63.
- Roberts, David. 2020. 'With Impeachment, America's Epistemic Crisis Has Arrived'. *Vox* (blog). 6 February 2020. <https://www.vox.com/policy-and-politics/2019/11/16/20964281/impeachment-hearings-trump-america-epistemic-crisis>.
- Rosenfeld, Sophia. 2019. *Democracy and Truth: A Short History*. Philadelphia: University of Pennsylvania Press.
- Ruddick, William. 1988. 'Social Self-Deception'. In *Perspectives on Self-Deception*, edited by Brian P. McLaughlin and Amelie Oksenberg-Rorty, 380–89. London: University of California Press.
- Solomon, Robert C. 2009. 'Self-Deception, and Self-Deception in Philosophy'. In *The Philosophy of Self-Deception*, edited by Clancy Martin, 15–36. Oxford: Oxford University Press.
- Stanley, Jason. 2018. *How Fascism Works: The Politics of Us and Them*. New York: Random House.
- Strickland, April A., Charles S. Taber, and Milton Lodge. 2011. 'Motivated Reasoning and Public Opinion'. *Journal of Health Politics, Policy and Law* 36 (6): 935–44.
- Taber, Charles S., and Milton Lodge. 2006. 'Motivated Skepticism in the Evaluation of Political Beliefs'. *American Journal of Political Science* 50 (3): 755–69.
- Tetlock, P. E., and A. Levi. 1982. 'Attribution Bias: On the Inconclusiveness of the Cognition-Motivation Debate'. *Journal of Experimental Social Psychology* 18: 68–88.
- Van Bavel, Jay J., and Andrea Pereira. 2018. 'The Partisan Brain: An Identity Based Model of Political Belief'. *Trends in Cognitive Sciences* 22 (3): 213–24.
- Wentura, D, and W Greve. 2003. 'Who Wants to Be Erudite? Everyone! Evidence for Automatic Adaptations of Trait Definition'. *Social Cognition* 22: 30–53.

- . 2005. 'Evidence for Self-Defensive Processes by Using a Sentence Priming Task'. *Self and Identity* 4: 193–211.
- Williams, Bernard. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Oxfordshire: Princeton University Press.
- Zajonc, Robert B. 1980. 'Feeling and Thinking: Preferences Need No Inferences'. *American Psychologist* 35: 151–75.
- Zvi Baron, Ilan. 2018. *How to Save Politics in a Post-Truth Era*. Manchester: Manchester University Press.

ⁱ This indirect account of how we might come to be self-deceived about our truthfulness via being self-deceived about first-order beliefs is set out in more detail at the beginning of section three.

ⁱⁱ For the sake of ease, the focus throughout shall be on the demands of truthfulness in relation to the acquisition of beliefs, though much of what will be said will apply to the maintenance and revision of existing beliefs also.

ⁱⁱⁱ For the purposes of this paper we shall take truthfulness to refer specifically to accuracy, though this is not intended to imply, contra Williams, that accuracy is a sufficient condition of truthfulness.

^{iv} Though that the external obstacles will have an inner representation and hence imply inner attitudes to how we address them means we ought not to draw the distinction between these and inner obstacles too starkly.

^v To be clear from the start, the 'Remainer' is intended as a specific example of a more general phenomenon. The example is intended to be illuminating; nothing hangs on the issue being Brexit.

^{vi} This is not to say that other explanations are not available to the Remainer which do not require him to purport a failure of accuracy to those who voted Leave, including those that revolve around the other virtue of truthfulness, sincerity. It is just, for the purposes of this scenario, that our Remainer finds this particular explanation more plausible.

^{vii} The account that shall be defended here is that the Remainer is self-deceived as to their own truthfulness, and, to pre-empt the discussion in the next section, to be self-deceived as to p is to believe that p is true. That is a necessary feature of self-deception insofar as a self-deceived belief is still a belief, and, as we have discussed in this section, belief is directly connected with truth. In that regard, in

attributing self-deception to others we can presume that they are committed to truthfulness, otherwise there would be no need for the self-deceit that obscures their epistemic failings from them. They could just hold p without any resulting cognitive dissonance.

^{viii} As Raphael Demos put it (1960, 588) in an early influential article: ‘Self-deception exists, I will say, when a person lies to himself, that is to say, persuades himself to believe what he knows is not so’.

^{ix} For a helpful overview of both these puzzles and the various philosophical responses to them see Baghramian and Nicholson 2013.

^x For influential accounts of self-deception that do understand it as intentional in nature see Pears 1984; Davidson 1986.

^{xi} To be clear, what follows can accommodate the possibility that self-deception can be intentional. The claim is merely that this is not the most common form that it takes, nor that it is the form in which being self-deceived about our own truthfulness is more plausible. Or, put differently, that the self-deception be intended is neither a necessary nor sufficient condition for self-deception.

^{xii} The literature on irrational belief formation or motivated reasoning is now vast. Important work that contributed to its emergence in the psychological literature include Kruglanski 1989; Tetlock and Levi 1982; Festinger 1957; Lord, Ross, and Lepper 1979; Kunda 1990; Pears 1984. For excellent examples of theories of motivated reasoning being applied to issues in political science see, for example, Druckman, Kuklinski, and Sigelman 2009; Taber and Lodge 2006; Druckman and Bolsen 2011; Pietryka and Boydston 2012; Bisgaard 2015; Enns and McAvoy 2012; McCabe 2016; Lodge and Taber 2013. The term ‘motivated reasoning’ will be used here to avoid the controversy as to whether the motivated belief is irrational.

^{xiii} Though this formulation is my own, it is firmly embedded in and in keeping with that strand of the literature that understands self-deception as a form of motivated reasoning, as pioneered in particular by the work of Alfred R. Mele (1987; 1997; 2001; 2003; 2004). It is worth saying that though this account is presented here primarily because it has, in my view, the virtue of being the best account of self-

deception, i.e. it is right, in an important sense the argument of this paper does not depend upon that being the case. The possibility that we can be self-deceived about our own truthfulness is, presumably, compatible with any account of self-deception. Where the particulars of this account are significant is in the explanation as to how particular beliefs can be selected for self-deception, and hence in justifying the claim that truthfulness might be something over which we are particularly prone to being self-deceived.

^{xiv} For Mele's own reflections on the role of emotions in self-deception see Mele 2003.

^{xv} Or, put differently, for it to be credible that someone is self-deceived in relation to any of these we would need to tell a quite specific story about why for them so much hangs on the answer to those questions.

^{xvi} The causal role that emotions can play in irrational belief formation is backed up by a now substantial body of psychological research that covers the variety of ways in which this can occur, including associating the emotions with less systematic thinking, less efficient processing skills, decreased reliance on direct evidence, decreased working memory capacity, and selective attention to evidence (for overviews see Izard, Kagan, and Zajonc 1984; Zajonc 1980; Niedenthal and Kitayama 1994; Dolan 2002).

^{xvii} Mele (2004) offers a plausible formal version of what this condition must be in his 'impartial observer test': 'If S is self-deceived in believing that p, and D is the collection of relevant data readily available to S, then if D were made readily available to S's impartial cognitive peers (including merely hypothetical people) and they were to engage in at least as much reflection on the issue as S does and at least a moderate amount of reflection, those who conclude that p is false would significantly outnumber those who conclude that p is true'.

^{xviii} As Baghramian and Nicholson note, this requires a violation of the Principle of Total Evidence, 'the epistemic norm enjoining us, when choosing among a set of mutually exclusive hypotheses, to give more credence to the one most strongly supported by all available relevant evidence' (2013, 1019).

^{xix} It is easy to imagine how there could be something of a feedback effect here: in wishing to protect our first-order beliefs from challenges we hold fast to our truthfulness, which in turn makes us more confident that we are truthful and hence that our first-order beliefs are indeed true, so we are less likely to countenance that they might be false, and so on. In those cases where people are self-deceived as to their truthfulness by virtue of being self-deceived about their first order beliefs we can see that there is a double motivation for defending one's truthfulness: the desire to be truthful but also the desire that those first order beliefs be true which motivated the self-deception in the first place. In that sense we might think that the route to being self-deceived about our truthfulness is much easier for those whose first order beliefs are also the result of self-deception.

^{xx} For a version of this history, purposefully retold as a means of reaffirming the Enlightenment notion of truth, see Rosenfeld 2019.