Is identity illusory?

Andreas L. Mogensen

Faculty of Philosophy, University of Oxford, Oxford, UK

Correspondence
Andreas L. Mogensen, Faculty of Philosophy, University of Oxford, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK.
Email: andreas.mogensen@philosophy.ox.ac.uk

Abstract
Certain of our traits are thought more central to who we are: they comprise our individual identity. What makes these traits privileged in this way? What accounts for their identity centrality? Although considerations of identity play a key role in many different areas of moral philosophy, I argue that we currently have no satisfactory account of the basis of identity centrality. Nor should we expect one. Rather, we should adopt an error theory: we should concede that there is nothing in reality corresponding to the perceived distinction between the central and peripheral traits of a person.

1 | INTRODUCTION

Any individual person exhibits an enormous range of traits and properties. You are so many feet and inches in height. Your eyes are a certain color. You have a view about whether God exists. You are 2.5 million light years from the Andromeda Galaxy. Within this set of properties, some are thought to be privileged. Certain traits are thought to be more central to who you are: they comprise your individual identity. For example, your religious beliefs would ordinarily be thought more determinative of who you are than your eye color.

Attempts to enumerate these supposedly central traits are likely to generate a heterogeneous and open-ended list of qualities. A given individual might be defined in part by her professional role, her religious beliefs, her degree of extraversion, her political orientation, her tastes and preferences, her nationality, her sexuality, and more besides. What makes these particular traits privileged: what accounts for their identity centrality? Why are they more integral to who she is than her belief that Bismarck unified Germany or the itch on her left shin?

In this paper, I will try to argue that we have no good answer to this question. Nor, I think, should we expect one. The failure of existing accounts leaves the nature of individual identity mysterious. Inductive inference should lead us to expect the failure of future efforts to define its basis. Furthermore, there exists a plausible debunking explanation that undermines our belief in individual identity.
Here is the plan. Section 2 will lay some necessary groundwork, clarifying the concept of individual identity at issue in this paper and illustrating its importance for a number of debates in moral philosophy. Section 3 criticizes extant accounts of the basis of identity centrality, showing that the theoretical proposals put forward by philosophers tend to conflict with certain core intuitions. Finally, Section 4 outlines a debunking explanation: namely, that the belief in individual identity is traceable to a pervasive but unreliable mental disposition, *psychological essentialism*. A brief summary and conclusion follow.

2 | GROUNDWORK

My aim in this section will be to clarify the nature of our subject matter and illustrate its importance for a range of topics in contemporary moral philosophy. I begin with some necessary points of clarification.

2.1 | Our subject matter

This paper is about individual identity understood in terms of those traits that define you as the person you are. Cognate terms include “the true self,” “deep self,” and “real self.” The term “individual identity” will be used to distinguish this sense of identity from the many others available in our conceptual repertoire. It will be helpful at the outset to make some of these distinctions clear.

Most obviously, individual identity must not be confused with *personal identity* as it is understood in contemporary metaphysics (Olson, 2010). Personal identity is an equivalence relation involving one and the same individual considered at different times. A person’s individual identity is not a relation, but a set of defining properties exemplified during a certain period of time. Speaking in terms of individual identity, it is perfectly intelligible to describe someone having undergone a profound personal transformation as being no longer the same person she once was, whereas this makes no sense as a claim about numerical identity.1

Individual identity should also be distinguished from certain notions of identity operative in social science and social theory. As it figures in “identity politics,” “identity” denotes either a socially salient category of group membership or the shared self-understanding through which such a group attains a sense of solidarity and distinctiveness.2 We should also distinguish our notion of individual identity from the concept of identity operative in Erik Erikson’s influential work on ego development, from which we derive the term “identity crisis” (Erikson, 1968). “Identity” in this sense corresponds roughly to “self-image” or “self-understanding”: your own (continually evolving)3 sense of what makes you. Our topic is not who you understand yourself to be, but who you really are. A person’s self-image might be out of sync with who she really is due to bad faith or simple ignorance.4

Lastly, individual identity should be distinguished from the phenomenon of *identification* as defined in Harry Frankfurt’s influential work in the philosophy of action.5 Roughly speaking, Frankfurt (1971, 1976, 1987, 2006) understands a person as identified with some attitude insofar as it receives reflexive endorsement via a corresponding wholehearted higher-order attitude. Frankfurt is clear that, in his view, being identified with one’s own mental states in this sense is a pervasive, mundane phenomenon, breaking down only in cases involving “obsessional thoughts,” “peculiar reckless impulses,” and “hot surges of anarchic emotion” (Frankfurt, 2006, pp. 8–9). The aim of Frankfurt’s theory is to distinguish these cases in which a person seems passive before the elements of her mind from those in which her attitudes are expressive of her mental activity.6

Frankfurt’s distinction between active and passive events in our mental lives is clearly not the same as our distinction between central and peripheral traits. Recall that being identified with one’s attitudes in Frankfurt’s sense is supposed to be the default condition: a person is identified with the vast majority of her attitudes, however, fleeting or trivial. For example, I am identified with my belief that the integral of $x^3$ is $0.25x^4 + c$. That is not just some idea
that occurs in me unbidden: it is the conclusion I formulate by thinking through the problem and one I wholeheartedly endorse. Nonetheless, few would suppose that this belief could be central to who I am.

In addition, it seems intelligible to suppose that a person's true self might include certain attitudes or feelings that are experienced as moving her against her own judgment. Consider Edith, a gay woman whose fundamentalist upbringing leads her to struggle against her own sexuality. We find it intuitive to describe Edith as trying to deny an important part of who she really is: a trait that is central to her identity but with which she does not identify (Arpaly, 2003, pp.14–16; Newman, Bloom, & Knobe, 2013).

2.2 | Its importance

Considerations of identity play a key role in many different areas of contemporary moral philosophy. The following is not intended as an exhaustive survey. I will focus on just four key issues: the ethics of biomedical enhancement; blame and responsibility; constructivist theories in meta-ethics; and the value of moral testimony.

The wide-ranging moral importance of individual identity plausibly reflects its intimate connection to the ethics of authenticity (Taylor, 1991). To a first approximation, authenticity is achieved when the way a person lives is expressive of her most centrally defining traits. Inauthenticity occurs when she fails to give expression to these traits. The key anxiety attached to the ideal of authenticity is that the conditions of modern life conspire to mask the true self beneath the demands of social conformity and the enticements of mass culture (Riesman, Glazer, & Denney, 1961/2001; Rousseau, 1782/2011). In spite of this perceived incongruity, authenticity is considered one of the constitutive ideals of modernity (Guignon, 2004; Taylor, 1989, 1991).

Considerations of authenticity have played a key role in recent debates on human enhancement (Juth, 2011). The specific type of enhancement at issue here is cosmetic psychopharmacology: the use of psychiatric drugs to bring about changes in mood and personality, allowing already healthy individuals to lead happier and more successful lives by becoming less shy, more confident, etc. (Kramer, 1993). Many find cosmetic psychopharmacology disturbing. In an influential paper, Elliott (1998) suggests that what disturbs us is the apparent inauthenticity involved in this kind of personal transformation: the pursuit of a new, enhanced personality represents a flight from the real you.7

Defenders of enhancement charge that Elliott’s concern rests on a mistaken conception of identity. DeGrazia (2000, 2005) argues that Elliott fails to appreciate the extent to which a person’s identity is determined by her own reflexive attitudes. Because of the authoritative role assigned to a person’s self-conception, DeGrazia concludes that if a person wholeheartedly desires to change some aspect of herself, she cannot meaningfully be accused of inauthenticity.8

Let us now turn to theories of blame and responsibility. It seems natural to suppose that the blameworthiness of a wrong act is proportional to the centrality of the attitudes that bring it about. Something like this is suggested in Hume’s well-known assertion that an individual cannot be blamed for his actions “where they proceed not from some cause in the characters and disposition of the person” and issue “from nothing in him, that is durable or constant” (Hume, 1739-40/2000, p. 262). Refining Hume’s suggestion, Sher (2005) argues that a person is blameable for an act insofar as its bad-making features can be traced to attitudes and dispositions that are central to her identity. A similar approach to praise and blame is suggested by Arpaly and Schroeder (1999) and by Sripada (2016).9

Considerations of identity also play a role in leading compatibilist responses to the problem of manipulation. Many compatibilist theories seem to have the counterintuitive implication that a person is morally responsible in cases where her behavior is brought about in a deterministic fashion via attitudes induced in her by extraneous forces, bypassing her own deliberative capacities (Pereboom, 2001). Mele (1995) and Haji (1998) argue that such agents lack responsibility not because their behavior is determined by factors beyond their control but because they act from attitudes that are not truly their own.10 Accordingly, they propose historical accounts of authenticity to supplement more familiar compatibilist criteria for appraisability.
Recent constructivist views in meta-ethics also involve prominent appeals to the normative significance of individual identity. Korsgaard (1996, 2009) assigns a central place to identity in her Kantian account of the sources of normativity. In her view, “Your reasons express your identity, your nature; your obligations spring from what that identity forbids.” (Korsgaard, 1996, p. 101) According to Korsgaard, it is only by drawing together normativity, obligation, and identity in this way that we can justify the extraordinary demands sometimes imposed on us by moral norms. Identity also has a role to play in Sharon Street’s Humean brand of constructivism. Where an agent’s normative attitudes conflict, Street insists that her reasons are determined by “whichever standpoint is most deeply hers” (Street, 2008, pp. 234–235).

Finally, I would like to highlight the role played by identity considerations in explaining our reluctance to accept moral testimony.

Although we can happily rely on the word of others when it comes to questions of about the geography of Spain or the dates of important historical events, many feel we cannot be quite so willing to defer on moral questions. In prior work, I have argued that the special importance we place on drawing our own moral conclusions reflects our attachment to authenticity as an ethical ideal (Mogensen, 2017). This ideal is said to place special demands on our moral beliefs, since our moral sensibility is considered most central to who we are.11

3 | WHAT IS THE BASIS OF IDENTITY CENTRALITY?

The foregoing section makes clear the importance of beliefs about individual identity for a wide range of philosophical issues. If identity is illusory, it seems that many otherwise plausible philosophical positions will be undermined and many key issues will have to be re-thought. The antecedent of this conditional is satisfied: there is no such thing as individual identity; there is nothing in reality corresponding to the perceived distinction between central and peripheral traits. My aim in the remainder of this paper will be to make the case for an error theory of this kind.

The first step in my argument will be to consider and reject various proposed answers to the question of what makes certain traits central to who we are. Note that my interest here will be in the basis of identity centrality, not its signs or symptoms. We may rely on various heuristics in deciding that certain traits belong to the true self without thereby picking up on any features that account for their belonging there. Consider an analogy. Most people would recognize a European robin (E. rubecula) based on its red breast. However, being red-breasted is not what makes a bird a robin. After all, robins only become red-breasted as adults.12 Similarly, we might have various criteria on which we rely in deciding that some trait is central to individual identity, whereas these criteria might not tell us why these traits have this kind of centrality.

I want to know what makes some trait central to individual identity. I do not presume that there is a single determinative property whose possession is both necessary and sufficient. There may be a number of different properties whose instantiation counts to some extent toward the centrality of a trait. This assumption will guide my discussion in the remainder of this section. I will consider various candidate properties, assessing the extent to which instantiation of these properties might account for the centrality of our supposedly core traits. I will argue that each proposal fails to accord with certain key intuitions about individual identity.

Before we move on, let me quickly highlight one issue that I’m not going to consider in what follows: namely, history. As suggested in the previous section, a trait that might otherwise seem to belong to the core of a person may instead be counted as an alien intrusion if it has the wrong history, arriving in the person’s mind via some deviant causal route involving psychosurgery, hypnosis, or remote manipulation. The question of how to mark off these deviant causal pathways is a difficult one, best avoided for present purposes. I assume that history is not the be-all and end-all of identity centrality. I will be considering which ahistorical properties might make some traits more central to who we are – assuming they have the right kind of history. We will begin by considering the relevance of durability.
3.1 Durability

Imagine Jane, a person in a constant state of flux. Her tastes bounce back and forth between various extremes. Her religious beliefs shift quickly from Anglicanism to Hinduism to Aztec revivalism. Even her character is fluid from 1 week to the next, flipping between introversion and extraversion. It may seem natural to say that Jane has no identity. Unless there is some enduring aspect of herself that she keeps hidden, she has no true self.

These reflections may suggest that an important part of what makes some trait central to a person's identity is its durability. According to Chandra Sripada, one characteristic feature of the attitudes comprising the agent's deep self is that they "tend to be stable and enduring features of a person's mental life." (Sripada, 2010, p. 165) Similarly, for Amélie Rorty and David Wong, one key dimension of identity centrality is the "degree to which it is difficult for a person to change the trait (which is often a function of its temporal persistence)" (Rorty & Wong, 1990, p. 20). These authors are not altogether clear on whether they understand durability as a basis for identity centrality or merely as a reliable indicator. I will argue that only the latter view is plausible.

First of all, we should draw an important distinction. We should distinguish de facto duration from durability. Durability is a dispositional quality. Roughly speaking, it is the disposition to resist change in the face of certain stressors. A trait of low durability might be held for considerable time if never challenged. A highly durable trait might vanish straightaway if subjected to extraordinary force. For this reason, it seems implausible that de facto duration matters to identity: the actual persistence of a trait depends as much on extraneous forces as it does on the nature of the trait and its bearer. Left in the clutches of the most efficient brain-washers, even our most central, defining qualities might vanish in the instant.

Persistence is relatively easy to define and measure. Durability is considerably harder to spell out. What are the relevant stressors in the face of which a durable trait endures? We obviously should not include advanced brain-washing techniques. What can we include? The difficulty of answering this question is compounded by the heterogeneity of the supposedly central traits. With respect to judgment-sensitive attitudes like beliefs and desires, it seems natural to suppose that their durability is a function of their capacity to resist revision in the face of reflection and deliberation. However, this criterion of durability makes little sense as applied to other traits, such as a person's sexuality or nationality.

Let us set aside this problem and focus on a person's judgment-sensitive attitudes and their stability in the face of deliberative appraisal. We can treat this as our best test-case for assessing the link between stability and identity. In so doing, I think we find that there is actually little to be said for the idea that durability counts toward identity centrality.

First, we should note that many of our most firmly held beliefs seem completely peripheral. Examples include my belief that that water is wet and that 2 + 2 = 4. My moral and religious beliefs are a good deal shakier, but would be said to be far more central to who I am.

More importantly, consider what happens when we compare these trivial but firmly held beliefs with similar beliefs that are more uncertain and less likely to be endorsed in future reflection: there is little temptation to say that the former are more central than the latter. For example, it seems peculiar to suggest that my belief that 2 + 2 = 4 is more central to my identity than my belief that (2 + 2)^4 = 256, although I can quite easily be led to retract the latter by failing to multiply correctly. Surely both beliefs are equally unconnected to who I really am, regardless of the fact that the former is more durable.

There remains the possibility that durability contributes to identity in the presence of certain further conditions that happen to go unsatisfied in the case of these arithmetic beliefs. If we have a strong initial impression that durability and identity are somehow connected, we might be tempted to hold out for this possibility.

However, we can do much to explain away the intuition that stability is a determinant of identity centrality. I've already noted that we might simply rely on durability as a heuristic for recognizing central traits. Furthermore, the vignette about Jane with which I opened this sub-section need not be described as depicting a person with no identity. Jane might equally well be said merely to lack a stable identity. She is a different person from 1 min to the next.
but that requires that she have some identity at each point in time. This instability might make it impossible to associate any determinate identity with the temporally extended person: Jane herself. So far as I can see, that is all that need be captured in saying that Jane has no identity.

3.2 | Uniqueness

It may seem plausible that the traits which define you should be to some extent distinctive or unique. The traits which make you who you are must be those which set you apart. In this vein, Charles Taylor writes: “Defining myself means finding what is significant in my difference from others.” (Taylor, 1991, pp. 35–36) A popular self-help book by Phil McGraw defines the authentic self as comprising those “strengths and values that are uniquely yours” (McGraw, 2002, p. 40).

Does the relative uniqueness of a trait really contribute toward its identity centrality? One problem for this view arises from the fact that we all have certain uniquely identifying properties that seem entirely peripheral. Your fingerprint may be unique to you. Still, you would not be considered a different person if your fingerprints changed. Suppose that although the number of hairs on your head is actually quite common, the number of hairs on your shin is completely unique. It seems absurd to insist that the number of hairs on your shin is therefore more central to who you are than the number of hairs on your head. Both are surely equally peripheral.

This issue need not prove too problematic for the authors I quoted earlier. Taylor does not say that defining myself means finding all the ways in which I differ from others. Rather, it means finding what is significant in how I differ from other people. The traits which belong to my identity must have some independent importance within the framework of my life. In the same vein, McGraw talks of uniqueness only in relation to people’s strengths and values. Let us therefore consider whether relative uniqueness might contribute to the centrality of those traits which are otherwise important or significant in these ways.

To do so, let us focus on the case of religion. A person’s religious beliefs are typically thought central to her identity. However, many religions have enormous numbers of adherents. There are more than a billion Roman Catholics. On its face, the mere popularity of a given religion is not determinative of the extent to which adherence to that religion belongs to the core of a person’s identity. Thus, it would be strikingly counterintuitive to suppose that Christianity becomes less central to the identity of ordinary Christians to the extent that missionaries succeed in finding new converts.

What is more, it seems plausible that we can explain away any tendency to associate uniqueness with identity centrality by positing that uniqueness is a recognitional heuristic tuned to a world in which inauthenticity is assumed to be a common occurrence. It is part of our ordinary thinking about identity that a person’s outward behavior need not manifest who they really are. Inauthenticity characteristically arises from conformity and imitation, or so we think. For this reason, we assume that inauthentic persons can be recognized by their lack of originality. Conversely, traits which are distinctive or unique are more likely to be indicative of who someone really is.13

3.3 | Depth and breadth of influence

If some element of a complicated system is described as “central” or “core,” we might reasonably expect that it plays a uniquely important role in controlling the overall behavior of the system. Similarly, we might expect that those traits that are most central to who we are attain that status at least in part in virtue of the depth and breadth of their influence.

There are different dimensions of depth and breadth to which we could appeal in developing this proposal. Consider hierarchical control. Our psychological states often exhibit a hierarchical ordering, with downstream beliefs and desires dependent on more fundamental upstream attitudes. Identity centrality might be partly a
matter of a trait's elevated position in this kind of chain-of-command. Consider how a person's religious commitments may be expected to echo throughout a vast number of downstream attitudes and behaviors. It may be natural to suppose that the number of these downstream elements provides an indication of how central religion is to the person.

Breadth of influence need not be understood purely in these terms. We might also be interested in the range of behaviors and attitudes affected. Some people behave very differently when at work as opposed to at home or with friends as opposed to their spouse. Someone whose behavior is too flexible in this way may be suspected of lacking an identity. This may suggest that part of what makes certain traits central to who we are is their capacity to shape our thoughts and activities across diverse domains of social behavior.

A number of authors put forward accounts of individual identity on which depth and breadth of influence figure prominently. For Arpaly and Schroeder, a "belief or desire is deep insofar as it is a powerful force in determining the actor's behavior, deeply held, deep-rooted." (Arpaly & Schroeder, 1999, p. 173) According to Sripada, the "deep self lies at the root of a person's motivational hierarchy and it exerts robust, stable, cross-situational effects in shaping long-term patterns of action." (Sripada, ms., p. 15 n. 30) For Rorty and Wong, key dimensions of trait centrality include "the extent to which other traits ... are dependent on it," "the extent to which a trait is exemplified across distinctive spheres," and the "extent to which it is dominant in conflicts with other traits" (Rorty & Wong, 1990, p. 20).

Should we accept that this kind of behavioral penetrance is an important contributor to identity centrality? I do not believe that we should.

One problem for this kind of view is that it seems too inclusive. It threatens to count as central certain traits that seem too basic and low-level to form part of our identity: the sort of core background processes that are necessary for getting by in day-to-day life.

Consider the network of doxastic dispositions that make up folk psychology. Reliance on folk psychology plays a central role in determining my behavior toward others throughout a broad range of social spheres. A vast range of my beliefs and desires thereby hinge on my use of folk psychology to predict how others act and think. Expectations generated by folk psychology are typically resistant to correction in light of conflicting evidence about human behavior. A roughly similar story can be told for folk physics, mutatis mutandis. In spite of this, it sounds odd to suggest that reliance on folk physics and/or folk psychology is central to who I am. They are more like the background against which my life takes shape than a proper part of it. In some sense, of course, they belong to the central elements of my mind. It is just not the sense of centrality that interests us.14

A converse problem for this sort of view arises from the possibility of inauthenticity. It is one of the central anxieties attached to our conception of individual identity that a person might fail to give expression to who she really is. Imagine Gemma, whose passion for philosophy is boxed in by the social expectation that she live as a traditional housewife. As she drives to pick up her children, philosophical questions sometimes rise in her mind and captivate her, but she learns to brush them off, assuming that no one will take her seriously. Or consider Charlotte, a Southern woman who moves to the East Coast and hides her roots to fit in: she affects a different accent, a new set of mannerisms, and various other behavioral strategies designed to mask her heritage. Eventually, this pretence becomes second nature.

An inauthentic existence of this kind would seem to be defined precisely by the failure of (some of) a person's central traits to achieve the kind of depth and breadth of influence that some authors believe to be partially constitutive of identity centrality. It is difficult to understand how a person could be said to mask an important facet of herself in this way if what makes a given trait central to who she is gets determined to a large extent by the degree to which it resonates throughout a wide range of her behaviors.

To accommodate the possibility of inauthenticity, we could say that although depth and breadth of influence contribute to identity centrality, they are not sufficiently important to rule out that a person's core traits could fail to gain a high degree of overall behavioral penetrance. But even the weaker claim that depth and breadth of influence contribute to some extent to identity centrality runs into trouble when considered in light of the ideal of authenticity.
This ideal may be understood as carrying the following normative implication. To the extent that some trait is
central to your identity during some period of time, you ought to make it the case that this trait gains depth and
breadth of influence with respect to your behavior in that period. Suppose, then, that the depth and breadth of influ-
ence of a trait during some period contribute to the identity centrality of the trait during that time. In that case, if
you give expression to the trait, it is more central to your identity; if you curtail its expression, it is less central.
Assuming that your reasons for giving expression to some part of your identity are stronger in proportion to its cen-
trality, this creates a baffling state of affairs in which your decision to increase or decrease the expression of some
trait appears to determine the strength of the obligation that you thereby satisfy or violate with respect to the
expression of that trait.

Put another way, we can say that when taken in conjunction with the ideal of authenticity, the suggestion that
depth and breadth of influence contributes to identity centrality seems to entail a violation of the principle that the
strength of one’s reasons for (or against) performing some action cannot be a function of whether the action is (or is
not) performed. This principle may be considered an implication of Erik Carlson’s principle of Normative Invariance,
according to which the normative status of an action does not depend on its performance (Carlson, 1995, p. 100).15

3.4 Reflexive attitudes

Earlier I emphasized that the sense of individual identity that interests us is conceptually distinct from the concept of
identity as self-image or self-understanding. We might nonetheless suspect that there is some important connection
between a person’s attitudes toward her traits and the extent to which they belong to her identity. There is clearly
something appealing about the suggestion that who you are is ultimately a matter of how you define yourself. A view
of this kind can seem liberating, though it may also be disquieting or destabilizing. A number of authors put forward
views of this kind, with different philosophers emphasizing different ways in which individuals might appropriate cer-
tain traits as more truly their own.

Some give authority to a person’s self-narrative. This view is especially associated with Marya Schechtman (1996). According to Schechtman, the centrality of a trait for a person is determined by its narrative sig-
nificance: the importance it holds within the self-told story of her life. The authority of a person’s self-narrative is
not absolute, however. Schechtman insists that we should not count as self-constituting those self-attributions that
are deeply out of touch with reality. In his work on the ethics of enhancement, DeGrazia follows Schechtman, telling
us that “one is the person described when we accept the bulk of one’s inner story but not those parts that are clearly
out of touch with reality.” (DeGrazia, 2005, p. 85).

A person’s self-narrative might give prominence to traits that she considers shameful or undesirable. On the
other hand, some philosophers emphasize a close tie between identity centrality and what we value in ourselves. For
Korsgaard, your identity is given by “a description under which you value yourself, a description under which you
find your life to be worth living and your actions to be worth undertaking.” (Korsgaard, 1996, p. 101) Similarly, for
Allen Buchanan, a person’s identity is “an understanding of who one is that embodies one’s most stable, highest pri-
ority values.” (Buchanan, 2011, p. 101)16 Sripada (2016) understands a person’s deep self as made up of her cares,
where part of what it is for some proattitude to represent a person’s care is that she is disposed to value its presence
and/or persistence.

A key problem for views of this kind arises from a possibility already noted at the end of Section 2.1. Intuitively,
a person can be in denial about who she really is. She may reject important parts of her identity, giving certain core
traits little or no place in her self-understanding. The example that we considered earlier was that of Edith, a person
raised in a fundamentalist household who is gay but repudiates her own sexuality. Many would say that her refusal
to accept herself as a gay woman is a form of inauthenticity.

You might suppose that this sort of case could be handled by the realism condition put forward by Schechtman.
On her view, a person’s narrative self-understanding loses its authority with respect to her identity insofar as that
life-story departs too far from the facts. A person who struggles against her sexuality may be expected to do so in light of certain deeply mistaken beliefs about the nature of homosexuality. Even so, this will not help with our problem. The realism condition is not a means for including certain traits within a person's identity in spite of her failure to acknowledge them: it is a tool for excluding certain traits that a person self-ascribes but should not.

Could we say that a person’s identity is not defined in terms of the reflexive attitudes she actually has (minus those excluded by the realism condition) but by those she would adopt under certain idealized epistemic conditions? These conditions might include full information and full deliberative rationality. We may expect that under such conditions, Edith would value her sexuality and assign it an important place in her narrative self-understanding.

However, this proposal falls prey to problems similar to those noted in Robert K. Shope’s classic paper on the conditional fallacy in philosophy (Shope, 1978). Certain of the traits that are important to who I am may be lost if I were fully informed and fully rational in my deliberations. One such trait might be my long-standing project of improving my cognitive habits so as to gradually approximate the epistemic ideal exemplified by a fully informed and fully rational deliberator. An ideal epistemic agent obviously would not adopt this kind of life-project and so would not include it in her self-definition. Alternatively, it might be a key part of who I am that I am hot-blooded and mercurial in a way that no ideally rational agent would be. Quite generally, my fallibility or irrationality may be important to making me the kind of person I am. For this reason, I think it would be a mistake to suppose that my identity can be defined in terms of the reflexive attitudes of my idealized counterpart.

A different way of accounting for the identity centrality of traits with which the agent fails to identify is suggested by Sripada (2016). According to Sripada, a person must have a disposition to value her possession of a given trait if that trait is to constitute a central part of her identity. This is compatible with her actually failing to value the trait, of course. Just as a fragile vase need not break to count as fragile, a person might be disposed to value certain of her desires without in fact valuing them.

It is not clear that this suggestion helps us to accommodate the example we have been working with so far. It sounds odd to say that a person whose religious beliefs lead her to repudiate her sexuality is disposed to value her sexuality. But here is one potential reply. In reminding us that a dispositional property need not be made manifest, Sripada emphasizes the capacity of dispositions to be defeated by masks and finks. A fragile vase might not break if struck due to the presence of protective packaging: its fragility is masked. We might claim that the person who struggles against her own sexuality is disposed to value being gay but this disposition is masked by the legacy of her upbringing. This suggestion may seem to accord with our earlier thought that Edith would value her sexuality and assign it an important place in her narrative self-understanding if she were placed under idealized epistemic conditions.

I am not convinced by this reply, however. Whereas it is widely accepted that dispositions can be masked by extrinsic factors, there is considerable controversy as to whether a disposition can be masked by something intrinsic to the individual or object. The possibility of intrinsic masking is endorsed by Randolph Clarke (2008, 2010) but ruled out by Choi (2005, 2012) and Handfield (2008). Someone who wants to insist that Edith has a masked disposition to value her sexuality is thus forced to take a highly controversial stand on the metaphysics of dispositions.

By contrast, we can deny that Edith is disposed to value her sexuality without taking such a stand: in particular, without insisting on the controversial claim that intrinsic masking is in principle impossible. It just seems quite intuitive to deny that Edith has the disposition in question. So far as I can tell, there is no good reason to overrule this intuition. Granted, it is true that Edith would value her sexuality if not for her fundamentalist upbringing. Still, that is no reason to suppose that she has a disposition to embrace her sexuality. It is equally true that birds would fall down in mid-air if they did not have wings, whereas there is no temptation to insist that birds have a disposition to fall out of the sky.  

Rather than try to accommodate the intuition that Edith is guilty of inauthenticity, some philosophers may try to reject it. The intuition may be said to belong to a partisan conception of authenticity. Levy (2007) argues that whereas some philosophers conceive of authenticity as involving fidelity to a pre-given set of defining characteristics, this assumption is not universal. According to Levy, there exists a rival tradition, exemplified in Sartre’s
philosophy, which rejects the assumption of a pre-given identity and emphasizes instead our total freedom. On this view, a person cannot be committed to accepting certain of her traits as central to who she is against the grain of her chosen self-definition.  

I think this objection rests on a natural but mistaken interpretation of Sartre’s philosophy. Sartre does in fact conceive of authenticity as a matter of fidelity to a pre-given, defining characteristic.

The reading suggested by Levy is, I concede, a very natural one. Sartre famously endorses the slogan “existence precedes essence,” and he tells us that a human being “to begin with ... is nothing. He will not be anything until later, and he will be what he makes of himself” (Sartre, 1946/2007, p. 22). These remarks clearly comport with the interpretation of Sartre suggested earlier. But they must be weighed against others in which Sartre seems, however paradoxically, to attribute some kind of fixed and inescapable nature to human beings.

Note, in particular, that Sartre understands each human being as defined most centrally by the fact of her freedom. He tells us that “there is no difference between the being of man and his being-free” (Sartre, 1943/1958, p. 25). This freedom includes the ability to unmake even our own freely chosen commitments: a prior resolution of mine binds me now, Sartre writes, only insofar as I "remake it ex nihilo and freely" (Sartre, 1943/1958, p. 33). We cannot help but exercise this radical freedom: we are condemned to be free. Nonetheless, we may refuse to acknowledge our existential predicament in our self-image or self-conception. This puts us in the condition of bad faith, characterized by Sartre as “a dissimulation of man’s full freedom of commitment” (Sartre, 1946/2007, p. 48). Authenticity instead means “adopting human reality as one’s own” (Sartre, 1983/1999, p. 113). We achieve authenticity by acknowledging our total freedom and living accordingly, without excuses. And this imperative of authenticity is not one that we can simply set aside. In the words of Jonathan Webber, Sartrean authenticity “is imperative for everyone irrespective of their projects” and “requires not only recognizing but also valuing the true structure of human existence” (Webber, 2018, p. 166).

The suggestion that Sartre attributes a fixed and inescapable nature to human beings may still seem jarring. I admit the formulation is chosen in part to be provocative. Nonetheless, it captures the truth. As Webber (2018, pp. 3–4) suggests, the key to escaping the appearance of paradox is to be careful in how we understand such terms as “essence” or “nature.” What Sartre rejects is a human essence understood in a distinctively teleological sense: an in-built nature oriented toward pre-defined goals. Understood as that which is centrally defining of us, freedom, as Sartre conceives it, does not carry with it a commitment to concrete goals. Unlike other traits that we might be tempted to think of as most central to who we are, acknowledging and valuing the fact of our freedom provides little determinate guidance for how to live. The authentic individual “is what he makes himself; that is all that can be said” (Sartre, 1948/1995, p. 137).

I conclude that Sartre does not propose a rival conception of authenticity on which a person is never committed to accepting certain traits as central to who she is in spite of her chosen self-definition. What is distinctive of Sartre’s conception is the nature of the trait that he treats as requiring acceptance and affirmation in order to realize our authenticity: in particular, the relative lack of concrete practical guidance for how to live that results from conceiving of authenticity in this way.

Of course, Sartre might not agree with the way in which many of us may be inclined to interpret Edith’s situation. While characterizing a person in denial of his homosexuality as “assuredly a man in bad faith” (Sartre, 1943/1958, p. 63), Sartre also depicts those who encourage us to view a person’s sexuality as making up a part of who they really are as in the grip of the very same failure mode (see Sartre, 1943/1958, pp. 63–65). Nonetheless, we could readily cook up an analogous case, similar to Edith’s, but re-written to fit the Sartrean emphasis on human freedom as our centrally defining trait. The underlying point would be the same.

There is a strong case, therefore, for thinking that our ordinary conception of identity and authenticity commits us to the possibility that a person may refuse to acknowledge or endorse certain of her centrally defining traits. Granting this, might there nonetheless be some more limited role to be played by a person’s reflexive attitudes in explaining the identity centrality of certain traits? For example, might it still be said that if Edith were to acknowledge her sexuality and come to define herself as a gay woman, then her sexuality would become more central to who she
is, at least to some extent? Obviously, it is true that being gay would become a more central part of her self-image or self-understanding, but that is not what is relevant here. Furthermore, accepting the conclusion seems to entail a violation of *Normative Invariance* when taken in conjunction with the intuitive assumption that authenticity makes demands on a person's self-image.

Authenticity requires that a person's identity gain expression in the cognitive and conative states that guide her through life. Assuming that a person's self-understanding and associated reflexive attitudes fall into that category, authenticity requires a person's self-image to reflect her identity. If identity centrality is in turn determined to some degree by the extent to which a trait gains recognition or endorsement within a person's self-understanding, we obtain the paradoxical state of affairs in which the strength of your reasons for including a certain trait within your self-definition seem to depend on whether or not you include that trait within your self-definition. We thus have another apparent violation of *Normative Invariance*. This suggests that we should reject the view that identity centrality depends to some degree on the extent to which a trait gains recognition or endorsement within a person's self-understanding.

To finish off this section, I will offer a back-handed concession to the view that a person's identity is determined by her reflexive attitudes. I think this view captures something of the way we talk about individual identity. However, I think that way of talking should not be taken literally.

Many of the traits which are said to belong to a person's identity are psychological in nature: beliefs, values, personality traits, etc. However, objective qualities may also be mentioned: being a philosopher, being a mother, etc. I think it is plausible that the conditions for saying that traits of the latter kind belong to a person's identity may depend on the person's reflexive attitudes. DeGrazia (2000) picks up on this in arguing that individual identity is a matter of self-definition. He asks us to contrast two inveterate smokers, Nina and Xena. Whereas Nina finds her smoking alien and out of character, Xena embraces it and delights in defying conventional good sense. As DeGrazia notes, it is natural to say that smoking is more definitive of Xena as a person than it is of Nina. And it is plausible that this is precisely because of a difference in their attitudes toward their respective habits.

I believe that when we describe a person's identity in terms of her objective qualities, we are really engaging in a form of metonymy. What we really think of as belonging to her identity is a set of subjective attitudes associated with the objective trait in question and not the trait itself. This can be seen by reflecting on the following kind of example. Imagine a person, Julian, of whom it seems natural to say that being a father is central to his identity. Now imagine that the following is true of Julian. Long before he became a father, Julian had determined that raising and caring for children would be a central part of his life. He knew exactly how many children he wanted to have and how he would parent them. When his children were born, he came to be just the kind of father he had always aspired to be.

A person for whom fatherhood is so important would naturally be described as having this trait at the center of his identity. At the same time, it seems odd to say that, when he became a father, Julian became a different person because only then did being a father become central to his identity. Prima facie, his identity did not change in this way. This cannot be true if the property of being a father is literally a central part of who he is now, since he did not have it previously. Our disposition to cite fatherhood as an important part of Julian's identity therefore seems best understood as a form of metonymy. We are really referring to the suite of attitudes toward fatherhood that he has had for a very long time.

A similar analysis applies to DeGrazia's example involving Nina and Xena. When we say that being a smoker belongs to Xena's identity, it is not the smoking itself but rather the set of attitudes toward smoking that we have in mind. This can be seen if we think of Xena as being like Julian. Suppose Xena consciously planned to defy conventional good sense by becoming an inveterate smoker: she is now living out that plan exactly as she envisioned it. It seems odd to suppose that her identity changed as a result of her carrying out what she had wholeheartedly intended to do all along. It must therefore be Xena's attitudes toward smoking that we really think of as belonging to her identity, not the smoking itself. Failure to mind this point may go some way to explaining the attraction of the view that the centrality of a trait is determined by the agent's reflexive attitudes.
4 | INDIVIDUAL IDENTITY AS A COGNITIVE ILLUSION

What could it be for some trait to be central to individual identity? My argument so far suggests that we may have no good answer to this question. Key proposals that philosophers have put forward seem implausible in one way or another.

Even if you agree with my argument in the previous section, you might think that my case so far does not offer convincing reasons for accepting an error theory. After all, it’s in the nature of philosophical reflection to confuse us about things we ordinarily take for granted. Philosophers love to pick apart one another’s views. We might not be especially surprised to discover that the theories which have been proposed so far to explain the basis of individual identity have a few issues.

However, my argument does not rest with the points noted in the previous section. In addition, I want to draw our attention to a plausible debunking explanation that allows us to see belief in individual identity as arising from an unreliable but pervasive psychological disposition. The disposition I have in mind is psychological essentialism. The suggestion that ordinary beliefs about identity may reflect essentialist thinking has been made before, but I intend to state the case more forcefully and comprehensively, as is needed if we are to debunk our intuitive belief in the existence of individual identity once and for all. This section will therefore describe the characteristic features of psychological essentialism, explain why I think beliefs about individual identity are symptomatic thereof, and then set out why I take this to constitute a debunking explanation.

4.1 | Psychological essentialism

The theory of psychological essentialism originates in research on concepts and categorization and is connected with the development of the theory paradigm of concepts (Medin & Ortony, 1989). A significant and still-growing corpus of psychological evidence indicates that we exhibit a broad-ranging disposition to conceptualize key features of the world in terms of essences: in particular, we are disposed to implicitly assume that what makes an individual the kind of individual it is depends on its possession of certain defining inner qualities, rather than any characteristic surface features. This disposition is observed across diverse cultures and appears to emerge in the developing child without explicit instruction (Gelman, 2003). Not only is this disposition cross-culturally robust, it is also robust in the face of conflicting evidence: scientific findings which contradict our essentialist assumptions are often inappropriately assimilated and misconstrued as evidence confirming these assumptions (Leslie, 2013, pp. 121–124).

Psychological essentialism is most notable as a pervasive feature of judgments about kind membership. Thus, people reliably assume that organisms fall under a common biological kind in virtue of having a similar essence, rather than because they share observable commonalities. For example, people assume that what makes a horse a horse is not the fact that it has hooves or neighs, but instead some internal set of properties shared with other horses. The essence is assumed to be unobservable and may in principle be completely masked by outward appearances. Thus, if a horse were transformed so that it looks and acts just like a zebra, people reliably say that it remains a horse nonetheless (Keil, 1989). As well as species, essentialism has been found to govern people’s judgments about many other categories including sex and race, thereby providing a rich vein of research for social psychologists interested in stereotyping and dehumanization (Prentice & Miller, 2007).

Although people readily categorize in terms of essences, they typically do so without any clear understanding of what these essences might be. Essences are projected automatically; only later do we form hypotheses about their nature. For this reason, Medin and Ortony (1989) describe people’s concepts as including an essence placeholder, capable of being filled in different ways.

People do, however, make certain characteristic assumptions about the causal powers of essences. For example, they assume that although the essence can in principle fail to be expressed (as in the example of a horse that looks and acts like a zebra), the essence ordinarily accounts for an individual’s observable traits and causes the
characteristic features associated with a given kind. Knowledge of essences therefore carries rich inductive potential. Essences are also assumed to be more durable than other properties and may even be conceived as immutable.25

Psychologists have proposed a number of additional assumptions as characteristic features of essentialist thinking, such as the assumption that kinds defined in terms of essences have sharp and unambiguous boundaries (see Gelman, 2003, pp. 21–4). However, these various assumptions about the properties of essences do not necessarily occur together and may in principle decompose into a number of distinguishable factors (Haslam, Rothschild, & Ernst, 2000). Essentialism is not all-or-nothing. The degree to which a given trait or category is essentialized may be indexed by the number of characteristic essentialist assumptions with which it is associated. Rather than a sharp dichotomy between essence and surface, Medin and Ortony propose “a continuum of centrality ranging from relatively inaccessible, deep properties to more accessible, surface ones.” (Medin & Ortony, 1989, p. 185).

4.2 | Belief in individual identity as a symptom of essentialist thinking

The suggestion, then, is that belief in individual identity is symptomatic of psychological essentialism. I will now explain why we should take this hypothesis seriously and why we should think it sufficiently plausible to throw belief in individual identity into doubt.

First, we can note that essentialist language is often employed to express claims about individual identity. For example, Frankfurt equates identity with “the essential nature of a person” (Frankfurt, 1993, p. 114). Similarly, Robert Paul Wolff writes that he does not simply happen to be a husband, a father, or a Jewish American raised in New York. Rather, he says, “I am essentially such a man.” (Wolff, 1976, p. 137).

I do not think it’s any coincidence what we can express claims about individual identity using the language of essences. The belief that certain traits are more central to who we are and others more peripheral seems on its face to straightforwardly embody the key essentialist assumption that although individuals have various recognizable surface qualities, they also have certain core properties which properly define them.

We might also note that key claims made by philosophers about the properties of our central traits map on to characteristic assumptions of essentialist thinking. As we have seen, a number of philosophers suppose that our core traits are typified by their durability. Similarly, a key feature of essentialist thinking is the assumption that essences are durable and likely to persist throughout observable changes. Many philosophers have supposed that our central traits exhibit depth and breadth of influence. In the same way, it is assumed that essences typically cause the observable properties of individuals and exhibit high inductive potential. The behavioral penetrance that we normally associate with our central traits can in principle break down, as in cases of inauthenticity. Similarly, in spite of their characteristic inductive potential, essences can in principle be masked by surface qualities, as in the example of a horse modified to behave and look exactly like a zebra. Lastly, the fact that we believe so confidently in the reality of individual identity in spite of lacking any plausible account of what identity centrality consists in strongly resembles the essentialist disposition to automatically project essences without any clear understanding of what those essences might be.

We need not rest content with this intuitive sense of fit, nor with the pronouncements of philosophers. Empirical evidence also suggests that psychological essentialism underlies judgments of identity centrality in the general population. In an effort to test essentialist beliefs about personality, Haslam, Bastian, and Bissett (2004) constructed an essentialism index designed to capture the extent to which different personality traits are associated with assumptions characteristic of essentialist thinking (inductive potential, durability, etc.). In comparing a trait’s score on this index against judgments of identity centrality made by undergraduate psychology students, they found that most of the variation in centrality judgments could be predicted by differences in the extent to which personality traits are essentialized.26

I believe the considerations rehearsed so far lend considerable weight to the hypothesis that beliefs about identity centrality are symptomatic of essentialist thinking. Some philosophers may nonetheless resist the suggestion that
their views about individual identity are manifestations of psychological essentialism. In particular, I have in mind those philosophers like DeGrazia who emphasize that our identity is not something fixed and can in principle be shaped by active self-creation. If we hold a view of individual identity according to which who we are is not some immutable natural fact, is it still plausible to suppose that we are dealing with a symptom of psychological essentialism?

I believe it is. It is true that naturalness and innateness are characteristic assumptions of essentialist thinking. However, as I have already noted, essentialism is not all-or-nothing and the full cluster of essentialist assumptions need not occur together. The view for which I am arguing does not require that beliefs about individual identity embody every hallmark of essentialist thinking. The claim that belief in the phenomenon of identity centrality is ultimately traceable to essentialism is quite consistent with the possibility that certain characteristic essentialist assumptions are overruled when philosophers explicitly theorize about individual identity. Views like DeGrazia’s may simply be less essentialist than others, while nonetheless having their origins in essentialist thinking.27

4.3 Psychological essentialism as a debunking explanation

Suppose we find it plausible that belief in individual identity is in fact symptomatic of psychological essentialism. Why should this help support an error theory about identity?

The hypothesis supports an error theory because scientific evidence reliably indicates that there do not exist essences corresponding to the projections of psychological essentialism.28 Consider the case of species. As noted earlier, people typically assume that what makes an organism a member of a species is its possession of a characteristic species essence. We might suppose that all pandas have some particular DNA sequence in virtue of which they count as pandas and in virtue of which they acquire a taste for bamboo and other recognizable surface properties. However, it is a matter of consensus in the philosophy of biology that species membership should not be defined in this way. Species membership is instead understood as grounded in relational properties, such as genealogical relatedness (Ereshefsky, 2010). Attempts to isolate genetic essences believed to underlie distinctions of race and sex similarly find our intuitions out of step with biological reality (Ainsworth, 2015; Kitcher, 2007).

Essentialism may even mislead us about the nature of chemical kinds. Consider the widely held assumption that water has a chemical essence specified by the compositional formula H2O. In fact, water in its liquid phase contains a range of ions, polymers, and combinations thereof, without which it would lack many of its characteristic surface properties, such as its high boiling point.29 The assumption that chemistry has discovered H2O to be the chemical essence of water is questionable: it is doubtful that chemistry has discovered any microstructural essence for water at all (Weisberg, 2006).

Given the documented unreliability of essentialist thinking, we ought to distrust beliefs which we have reason to believe reflect the influence of psychological essentialism. In particular, we ought to distrust the sense of intuitive obviousness with which such beliefs are apt to strike us (Leslie, 2013, pp. 108–109). Thus, if belief in individual identity is symptomatic of essentialist thinking, this should cast significant doubt on the reality of our intuitive distinction between central and peripheral traits. At the very least, this explanation allows us to see why we could be expected to believe in the phenomenon of identity centrality even if individual identity is ultimately illusory. We are reliably disposed to believe in essences, even where none exist and where no evidence exists to suggest they do. Supposing that we in fact simply have the traits we do and nothing more, we should expect that creatures like us would impose an illusory ordering on these traits, arranging them along a continuum from a definitional core to an outward, front-facing periphery.30

5 Conclusion

I have argued for an error theory with respect to individual identity. On my view, there is nothing in reality corresponding to our perceived distinction between the central and peripheral traits of a person. I do not deny that
some of our traits are distinguished in terms of their durability, their uniqueness, their breadth and depth of influence, or their importance in our self-definition. However, none of these properties serves as a plausible basis for identity centrality. It is unclear what could. At the same time, it seems plausible that belief in individual identity is symptomatic of psychological essentialism: a pervasive but unreliable disposition to conceptualize individuals in terms of a definitional core lying beneath a peripheral surface. Given the documented unreliability of essentialism as a psychological disposition, it would be unsurprising to discover that there is in fact nothing corresponding to our intuitive conception of individual identity. We have many different properties, but none are really any more or less central to who we are.

The natural inference to draw from this conclusion is that authenticity is an empty ideal. Just as those of us who reject the reality of reincarnation should deny that release from samsara is a goal worth pursuing, so we ought now to deny that authenticity is something at which we ought to aim. As a result, the ideal of authenticity cannot play the roles it has been asked to play by philosophers. For example, it cannot serve as a basis for ethical criticism of human enhancement, nor for an imperative to develop one’s own moral views as opposed to acquiring moral beliefs by deference to testimony. There may, of course, be virtues worth pursuing in the neighborhood of the ideal of authenticity, such as plain honesty. But authenticity itself will have to go.31

ORCID
Andreas L. Mogensen https://orcid.org/0000-0001-6239-8710

ENDNOTES
1 For similar reasons, the view for which I argue in this paper bears only a tenuous relationship to the Buddhist doctrine that "the self" does not exist. On the distinction between individual identity and "the self," so understood, see Siderits (2007, pp. 32–34).
2 And more besides: see Brubaker and Cooper (2000).
3 Erikson objects to the equation of identity with "terms such as self-conception, self-imagery, or self-esteem" (Erikson, 1968, p. 23) because he thinks of these as connoting something "static and unchangeable" (p. 24), thereby obscuring his central developmental concerns.
4 This is not to rule out that there may be some important connection between self-image and individual identity. See Section 3.3.
5 The importance of recognizing this distinction in interpreting Frankfurt's work is noted by Scanlon (2002) and Velleman (2002).
6 Frankfurt does address the topic of individual identity in his work on caring, love, and volitional necessity. See Frankfurt (1982, 1988a, 1993, 1999, 2006).
7 See also Elliott (2003).
8 See also Levy (2011). For valuable commentary on this debate see Buchanan (2011), Erler (2012), and Parens (2005).
9 See also Faraci and Shoemaker (2010, 2014). In this vein we might also consider Wolf's discussion of "real self" theories of responsibility (Wolf, 1990). However, Wolf's conception of the "real self" tends to blur the lines between identity centrality and Frankfurtian identification.
10 See also Haji and Cuypers (2008), Mele (2006), Sripada (2012).
11 See Riis, Simmons, and Goodwin (2008), Strohminger and Nichols (2014), Taylor (1989).
12 What makes a bird a robin, roughly speaking, is its place within a particular genealogical lineage. On the basis of species membership, see Ereshefsky (2010).
13 In this vein, it's notable that McGraw defines the authentic self as comprising those "strengths and values that are uniquely yours and need expression, versus what you have been programmed to believe that you are 'supposed' to be and do." (McGraw, 2002, p. 40. My emphasis.)
14 This might be taken to suggest that those who think of depth and breadth of influence as definitive of individual identity may simply be confused about the relevant notion of centrality.
15 Strictly speaking, Carlson's principle mentions only "moral status". The generalized principle set out above is due to Bykvist (2007). Bykvist argues plausibly that Normative Invariance should not be considered an inviolable criterion of
theory choice. I agree that a normative position violating this criterion could in principle be acceptable all-things-considered. Nonetheless, violating **Normative Invariance** counts strongly against the plausibility of a theory.

16 Paralleling Schechtman and DeGrazia, Buchanan wants to exclude from this picture any elements of a person’s self-conception that are “grounded in grossly false beliefs or in a ‘script’ that one has internalized without critical reflection.” (Buchanan, 2011, p. 102).

17 See Choi (2012) for discussion of this example.

18 Many other authors postulate this kind of dichotomy between Romantic and Existentialist conceptions of authenticity: for example, Varga (2012).

19 This, at least, is the view of the early Sartre. Webber (2018) argues that by the time he would come to write his biography of Jean Genet in 1952, Sartre had come to agree with Beauvoir’s view that a person’s prior commitments acquire a weight or inertia of their own and are not sustained only by the agent’s continually renewed free endorsement.

20 Strictly speaking, this is hyperbole, as Sartre believes that authenticity carries with it some substantive ends, such as a commitment to “w[illing] the freedom of others.” (Sartre, 1946/2007, p. 49) Whether he has good arguments for that conclusion is a different matter. See Poellner (2015) and Webber (2018, chap. 9).

21 Exactly how Sartre means for us to understand the bad faith of the person in denial of his homosexuality is a matter of scholarly controversy. For contrasting views, see McCulloch (1994, pp. 61–62), Webber (2009, pp. 20–21, 80–84).

22 On authenticity and self-knowledge, see Feldman and Hazlett (2013).

23 This metonymic interpretation may well also be applicable to some claims about identity that involve psychological traits.

24 See Appiah (1996, p. 126), Leslie (2013, p. 112).

25 In connection with this point, it is perhaps worth emphasizing that essences so understood need not be essential properties as the latter are understood in metaphysics (Robertson & Atkins, 2016). To my knowledge, there is no assumption that essentialist thinking involves the belief that there is no metaphysically possible world in which an individual has a different essence.

26 This finding is correlational, leaving open the possibility that essentialist beliefs are explained by judgments of identity centrality, rather than vice versa. However, given the status of psychological essentialism as a pervasive and deep-rooted mental disposition, I take the converse to be the more plausible causal hypothesis. Another possibility is that some third factor explains both judgments of identity centrality and the extent to which personality traits are essentialized.

27 Compare Leslie (2013, pp. 115–116).

28 See Leslie’s (2013) authoritative survey, to which my discussion in the following is indebted.

29 See Leslie (2013, pp. 144–146) for further discussion and references.

30 Some readers might suspect that all I’ve managed in this paper is to recapitulate a well-known postmodernist view about the insubstantiality of the human person. Proponents of this view like Kenneth Gergen tell us that “there is no individual essence” (Gergen, 1991, p. 139). However, Gergen’s view rests on the empirical claim that individual behaviour is contextually variable and situationally determined to such an extent that we ought to deny the existence of stable, broad behavioural dispositions. This is really little more than a restatement of the situationist critique of personality psychology (Mischel, 1968), albeit in an extreme form. Unlike Gergen, I do not deny the reality of personality traits; I take it their existence is well-supported by the available evidence (Fleeson & Noftle, 2008). For reasons made clear in Section 3, the existence of stable, broad behavioural dispositions is insufficient to guarantee the reality of individual identity.

31 For helpful feedback and discussion of the ideas in this paper, I would like to thank Alexandre Erler and Stephen Mulhall, as well as an anonymous referee as *EJP*.

**REFERENCES**

Ainsworth, C. (2015). Sex redefined. *Nature*, 518, 288–291.

Appiah, K. A. (1996). *Race, culture, identity: Misunderstood connections The tanner lectures on human values* (Vol. 17, pp. 51–136).

Arpaly, N. (2003). *Unprincipled virtue: An inquiry into moral agency*. Oxford, England: Oxford University Press.

Arpaly, N., & Schroeder, T. (1999). Praise, blame and the whole self. *Philosophical Studies*, 93, 161–188.

Brubaker, R., & Cooper, F. (2000). Beyond “identity”. *Theory and Society*, 29, 1–47.

Buchanan, A. E. (2011). *Beyond humanity? The ethics of biomedical enhancement*. Oxford, England: Oxford University Press.

Bykvist, K. (2007). Violations of normative invariance: Some thoughts on shifty oughts. *Theoria*, 73, 98–120.
Kramer, P. D. (1993). Listening to prozac: A psychiatrist explores antidepressant drugs and the remaking of the self. London, England: Fourth Estate.

Leslie, S. J. (2013). Essence and natural kinds: When science meets preschooler intuition Oxford studies in epistemology (Vol. 4, pp. 108–166).

Levy, N. (2007). Neuroethics: Challenges for the 21st century. Cambridge, UK: Cambridge University Press.

Levy, N. (2011). Enhancing authenticity. Journal of Applied Philosophy, 28, 308–318.

McCulloch, G. (1994). Using Sartre: An analytical introduction to early Sartrean themes. London, England: Routledge.

McGraw, P. (2002). Self matters: Creating your life from the inside out. London, England: Simon & Schuster.

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 179–195). Cambridge, UK: Cambridge University Press.

Mele, A. R. (1995). Autonomous agents: From self-control to autonomy. Oxford, England: Oxford University Press.

Mele, A. R. (2006). Free will and luck. Oxford, England: Oxford University Press.

Mischel, W. (1968). Personality and assessment. New York, NY: John Wiley & Sons.

Mogensen, A. L. (2017). Moral testimony pessimism and the uncertain value of authenticity. Philosophy and Phenomenological Research, 95, 261–284.

Newman, G. E., Bloom, P., & Knobe, J. (2013). Value judgments and the true self. Personality and Social Psychology Bulletin, 40, 203–216.

Olson, E. (2010). Personal identity. In Zalta (Ed.), The stanford encyclopedia of philosophy. Retrieved from http://plato.stanford.edu/entries/identity-personal/.

Parens, E. (2005). Authenticity and ambivalence: Toward understanding the enhancement debate. The Hastings Center Report, 35, 34–41.

Pereboom, D. (2001). Living without free will. Cambridge, UK: Cambridge University Press.

Poellner, P. (2015). Early Sartre on freedom and ethics. European Journal of Philosophy, 23, 221–247.

Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. Current Directions in Psychological Science, 16, 202–206.

Riesman, D., Glazer, N., & Denney, R. (1961/2001). The lonely crowd. New Haven, CA: Yale University Press.

Riis, J., Simmons, J. P., & Goodwin, G. P. (2008). Preferences for enhancement pharmaceuticals: The reluctance to enhance fundamental traits. Journal of Consumer Research, 35, 495–508.

Robertson, T., & Atkins, P. (2016). Essential vs. accidental properties. In Zalta (Ed.), The stanford encyclopedia of philosophy. Retrieved from https://plato.stanford.edu/archives/sum2016/entries/essential-accidental/.

Rorty, A. O., & Wong, D. (1990). Aspects of identity and agency. In O. Flanagan & A. O. Rorty (Eds.), Rousseau, J. J. (1782/2011).

Rousseau, J. J. (1782/2011). War diaries: Notebooks from a phoney war, 1939-40, transl. Hoare. London, England: Verso.

Rorty, A. O., & Wong, D. (1990). Aspects of identity and agency. In O. Flanagan & A. O. Rorty (Eds.), Rousseau, J. J. (1782/2011).

Scanlon, T. M. (2002). Reasons and passions. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 179–195). Cambridge, UK: Cambridge University Press.

Schechtman, M. (1996). The constitution of selves. Ithaca, NY: Cornell University Press.

Sher, G. (2005). In praise of blame. Oxford, England: Oxford University Press.

Shope, R. K. (1978). The conditional falling in contemporary philosophy. Journal of Philosophy, 75, 397–413.

Siderits, M. (2007). Buddhism as philosophy: An introduction. New York, NY: Ashgate.

Sripada, C. S. (2010). The Deep Self Model and asymmetries in folk judgments about intentional action. Philosophical Studies, 151, 159–176.

Sripada, C. S. (2012). What makes a manipulated agent unfree? Philosophy and Phenomenological Research, 85, 563–593.

Sripada, C. S. (2016). Self-expression: A deep self theory of moral responsibility. Philosophical Studies, 173, 1203–1232.

Sripada, C. S. (ms). At the center of agency, the deep self. Unpublished manuscript. Retrieved from https://umich.box.com/s/vlkn6s4ggnc7uefft2

Street, S. (2008). Constructivism about reasons. In R. Shafer-Landau (Ed.), Oxford studies in meta-ethics (Vol. 3, pp. 207–245). Oxford University Press.

Strohminger, N., & Nichols, S. (2014). The essential moral self. Cognition, 131, 159–171.

Taylor, C. (1989). Sources of the self: The making of the modern identity. Cambridge, MA: Harvard University Press.

Taylor, C. (1991). The ethics of authenticity. Cambridge, MA: Harvard University Press.

Varga, S. (2012). Authenticity as an ethical ideal. Abingdon, England: Routledge.
Velleman, J. D. (2002). Identification and identity. In S. Buss & L. Overton (Eds.), Contours of agency: Essays on themes from Harry Frankfurt (pp. 91–123). Cambridge, MA: MIT Press.

Webber, J. (2009). The existentialism of Jean-Paul Sartre. Abingdon, England: Routledge.

Webber, J. (2018). Rethinking existentialism. Oxford, England: Oxford University Press.

Weisberg, M. (2006). Water is not H2O. In D. Baird, E. Scerri, & L. McIntyre (Eds.), Philosophy of chemistry: Synthesis of a new discipline (pp. 337–346). Dordrecht, the Netherlands: Springer.

Wolf, S. (1990). The real self view. In J. M. Fischer & M. Ravizza (Eds.), Perspectives on moral responsibility (pp. 151–169). Ithaca, NY: Cornell University Press.

Wolff, R. P. (1976). There's nobody here but us persons. In C. C. Gould & M. W. Wartofsky (Eds.), Women and philosophy: Toward a theory of liberation (pp. 128–144). New York, NY: G. P. Putnam's Sons.

How to cite this article: Mogensen AL. Is identity illusory? Eur J Philos. 2021;29:55–73. https://doi.org/10.1111/ejop.12552