

ARTICLE



Natural belief in persistent selves

Mark Collier

Department of Philosophy, Professor of Philosophy University of Minnesota, Morris, MN, USA

ABSTRACT

In “Of Personal Identity”, Hume attempts to understand why we ordinarily believe in persistent selves. He proposes that this ontological commitment depends on illusions and fictions: the imagination tricks us into supposing that an unchanging core self remains static through the flux and change of experience. Recent work in cognitive science provides a good deal of support for Hume’s hypothesis that common beliefs about the self are founded on psychological biases rather than rational insight or evidence. We naturally believe in personal persistence, according to this emerging research, because we are prone to categorize the world in terms of hidden essences and structure our lives in terms of whole life stories.

ARTICLE HISTORY

Received 18 July 2019
Accepted 20 July 2021

KEYWORDS

Hume; personal identity; natural beliefs; psychological essentialism; narrativity; buddhism

1. Introduction

In section 1.4.6 of the *Treatise*, Hume adopts an experimental approach to the study of personal identity.¹ His main concern is to understand why it is so natural for human beings to believe in the continued existence of selves. We ordinarily suppose that we remain one and the same person over stretches of time. But what principles of human nature are responsible for this widespread belief? Is ontological commitment to persistent selves founded on reason or evidence? Or is it based on illusions and fictions of the imagination?

Hume offers a number of empirical claims about the psychological mechanisms that are responsible for common beliefs about personal persistence. We must evaluate the status of these proposals, therefore, in light of our best available theories and evidence. How well does Hume’s experimental account fare when evaluated from a cognitive science perspective? Does it teach us anything of lasting importance? Or is it merely of historical interest?

This paper maintains that cognitive science research offers *prima facie* support for Hume's explanation of everyday belief in persistent selves. Recent work on essential and narrative selves demonstrates that human beings have a natural tendency to suppose, contrary to the perpetual flux of experience, that they extend over time as immutable core selves or whole life stories. We are disposed to believe that there is greater coherence in our lives than we actually perceive, in other words, and that the successive moments of our personal past are united through narrative connections or hidden self-kernels.

Cognitive science research indicates that ontological commitment to persistent selves is based on universal propensities of the mind, although cultural factors influence whether individuals adopt essentialist or narrativist strategies to accommodate change and maintain a sense of identity. The fact that everyday belief in personal persistence is based on general principles of human nature serves to explain, in turn, why this conviction is so resistant to philosophical reflection. Everyday belief in persistent selves cannot be dislodged by skeptical arguments because it is not founded on reason or evidence; rather, it is a *natural belief* as unavoidable and indispensable to human beings as breathing.

1.1 Hume's experimental approach to personal identity

Hume begins his investigations with a descriptive account of the content of our everyday beliefs about selves. We ordinarily think that we remain *perfectly identical* over time in the sense that we “continue invariably the same, thro' the whole course of our lives” (T 1.4.6.2; SBN, p. 251). Let us capture this commitment in the following terms, with letters representing impressions.

(1) AAAAAAAAAAAAAA

We do not only suppose that we continue to exist as numerically one and the same person over long stretches of time, on this proposal, but we also commonly believe that we remain qualitatively unchanged. We ordinarily imagine that there are unchanging cores to selves – some kind of pearls or kernels – that remain fixed and invariant over time.²

What we commonly believe about ourselves, however, does not correspond to experience. We ordinarily suppose that we are acquainted with statically persistent selves; but what we actually encounter is a dynamic succession of perceptions in a state of “perpetual flux and movement” (T 1.4.6.4; SBN, p. 252):

(2) AAABBBCCDDDEEE

Phenomenological consciousness does not reveal any mental states that are “constant and invariable” and that stand still over long stretches of time (T 1.4.6.2; SBN, p. 252). Everyday belief in the perfect identity of selves, therefore, goes beyond the available evidence.

One might be tempted to respond that Hume is searching in the wrong place for a statically persistent self. We might not be able to discover any immutable pearls or kernels when surveying the fluctuating contents of experience. But perhaps it is the *subject* of these experiences which remains constant and invariable over time; it is the *owner* or *bearer* of these perceptions, in other words, which resists change.

Hume denies that we are phenomenologically aware of ourselves, however, as something that exists independently of perceptions. We are not acquainted with ourselves at any moment of time, according to Hume, let alone as temporally extended entities.

For my part, when I enter intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch myself at any time without a perception, and never can observe anything but the perception. (T 1.4.6.3; SBN, p. 252).

The subject of experience is not something which appears in conscious awareness, in other words, as something over and above the features of perceptual states. There is, to put it succinctly, no “phenomenal I” (Prinz, 2012).

Why do we ordinarily believe in statically persistent selves, then, if this commitment goes beyond the available evidence? Hume’s positive thesis in T 1.4.6 is that common beliefs about diachronic selves depend on biases and propensities of the imagination. We come to believe that selves are perfectly identical because the mind processes the changing sequences illustrated in (2) as smoothly and effortlessly as it does the constant and invariable series depicted in (1), and as result, we are naturally disposed to confound these different patterns and inadvertently substitute one for the other in our “common way of thinking” (T 1.4.6.6; SBN, p. 254).³ Belief in statically persistent selves is illusory, then, in the sense that phenomenal experience is incorrectly interpreted by the mind. We are really only acquainted with a series of changing perceptions, but the imagination tricks us into viewing these contents “as if” they were constant and invariable (*ibid.*).

Careful reflection and philosophy teaches us, according to Hume, that perceptions are fluctuating rather than static. But it is psychologically impossible to completely shake or extirpate this illusion of the imagination.⁴

Our propensity to this mistake is so great . . . that we fall into it before we are aware, and tho’ we incessantly correct ourselves by reflection, and return to a more accurate method of thinking, yet we cannot long sustain our philosophy, or take off this bias from the imagination (T 1.4.6.6.; SBN, p. 254).

Philosophers continue to assent to the existence of fixed and immutable selves, therefore, while at the same time acknowledging these commitments to be empirically vacuous. Unable to abandon their natural beliefs in statically persistent selves, they turn to confabulation and invent palliative “fictions” in order to “justify” these implicit convictions (T 1.4.6.6; SBN, p. 254).

Philosophical belief in immutable diachronic selves is *fictitious* in the sense that it is an artificial creation whose sole purpose is to make us feel better. Philosophers do not posit these fixed entities in response to evidence or on the basis of reasons or arguments; rather, they do so because it enables them to avoid the uncomfortable collision between their intuitive and reflective beliefs. Unable to rationally justify their naïve commitment to statically persistent selves, they “feign” the existence of immortal souls or thinking substances in order to “disguise the variation” in their conscious experiences (T 1.4.6.6; SBN, p. 254).

Hume offers a similar account of belief in the synchronic existence of selves. We ordinarily suppose that selves, at one moment of time, are simple entities which lack constituent parts. But what we actually perceive are complex bundles of co-occurring qualities. Ontological commitment to atomic selves is once again based on a confusion of the imagination. A simple and indivisible entity “operates upon the imagination after much the same manner” as a collection of closely related parts (T 1.4.7.22; SBN, p. 263). We are unwittingly apt to conflate these similar sequences, therefore, and substitute one for the other in our thought. Philosophers who carefully reflect on the matter recognize that these attributions are mistaken, but are unable to loosen the grip of this illusion; in order to support their natural conviction that synchronic selves are simple and indivisible, they are disposed to “feign” an unperceived “center” which supplies the missing sense of unity (T 1.4.6.22; SBN, p. 263).

Hume does not dogmatically assert that philosophical beliefs about selves are erroneous in the sense that they fail to correspond to objective states of affairs; rather, they are empirically false in that they fail to correspond to the available evidence. It is conceivable and thus epistemically possible that there are perfectly identical selves; there is nothing contradictory, as far as we can tell, about the notion of simple, indestructible, and unchanging self-kernels. But it is also possible that selves are momentary or short-lived entities. The crucial point is that the available evidence does not allow us to rationally decide between these alternatives. Common beliefs about personal identity are derived, then, from principles of imagination rather than reason.

2. Hume's experiments

Hume maintains that ordinary beliefs about personal identity are based on a fundamental attribution error: we are psychologically disposed to conflate fluctuating and invariant perceptions. Given that these mistaken ascriptions are based on natural tendencies of the mind, we would expect to find similar types of errors in our everyday thinking about persistence across a variety of object categories or domains. In order to test whether this is so, Hume conducts a series of experiments that survey our common attitudes about the diachronic identity of material objects, artifacts, and biological organisms. The crucial question is whether we are prone to attribute perfect identity and static persistence to objects in each of these cases.

Hume's first "experiment" asks us to consider masses of matter, such as rocks, that undergo a "very small or inconsiderable" addition or subtraction to their component parts (T 1.4.6.8; SBN, p. 255). This is a real alteration, even if a slight one. But Hume proposes that this is not how people would ordinarily respond to this vignette. We are psychologically disposed to *ignore* such negligible changes and attribute a constant and invariable existence to the rock. A bias of the imagination leads us to "scarce perceive" the transformation of its sensible qualities: we confound the observation of minor changes with the perception of something static and constant because the mind moves smoothly and easily over both types of sequences (T 1.4.6.8; SBN, p. 256).

This psychological explanation can be further "confirm'd", according to Hume, by conducting a second experiment (T 1.4.6.10; SBN, p. 256). Consider a material object that undergoes major changes, but it does so in a *gradual* manner. Hume maintains that we would succumb to the illusion of viewing this object as if it were fixed and immutable. The imagination once again moves easily across these perceptual sequences and makes us feel as though we are perceiving a qualitatively constant and invariable (AAAAAAAAAA) entity. Indeed, we often make this mistake in everyday life: one fails to notice incremental changes in the height of children, for example, or the slow and steady erosion of rock formations.

Hume's third experiment involves a case where small and gradual changes eventually lead to a radical transformation. He asks us to consider the famous case of Plutarch's Ship of Theseus, whose wooden planks are successively repaired such that, over a period of many years, a considerable number of them have been replaced (the sailors presumably cannot replace planks that are submerged beneath the water). Hume maintains that we would intuitively judge that the ship remains qualitatively unchanged despite the fact that a majority of its parts have been altered. This erroneous attribution can only be explained, therefore, in terms of illusions of the imagination. Even though a large number of the planks are replaced over

time, their “purpose” remains exactly the same, and this fact is sufficient to induce the imagination to ascribe a fixed and immutable persistence to the ship (T 1.4.6.11; SBN, p. 257).

Hume’s fourth and final experiment involves biological organisms such as vegetables and animals. Suppose there is a young sapling that eventually grows into a mature oak tree. This is a case where the parts of an object are “totally replaced”: the older tree will not contain any of the material particles originally present in the sapling (T 1.4.6.12; SBN, p. 257). The formal qualities of the tree will also have been transformed: it is now much larger, for example, and has a notably different shape. But we would nevertheless be inclined to believe, according to Hume, that the tree remains perfectly identical over time. This phenomenon can only be explained, therefore, in terms of biases and propensities of the imagination. The tree not only retains the same purpose or function over time, as is the case with the ship, but it does so because of the reciprocal interactions of each of its parts (T 1.4.6.12; SBN, p. 257). These causal relations prompt the imagination to feel an easy transition when surveying this changing object and confuse it with the perception of something that statically persists over time.⁵

Hume maintains that these four experiments demonstrate the explanatory power of his imagination-based account. His psychological explanation provides a unified account of our everyday intuitions about persistence. Indeed, Hume claims to have discovered a law-like generalization: in every instance where human beings observe a succession of closely related perceptions (AABBCCDDEE), they are disposed to make attributions of static persistence (AAAAAAAAAA). This pattern is robust across a wide range of cases. Hume cannot be fairly charged with offering an *ad hoc* explanation of everyday judgments of personal persistence, then, because his imagination-based hypothesis also covers ascriptions of diachronic identity to rocks, ships, vegetables and animals.

These experiments also serve a rhetorical function. One might be willing to accept that attributing perfect identity to mundane material objects involve fictions of the imagination. But one might be naturally inclined to refrain from endorsing this type of explanation with regard to something as privileged and cherished as ourselves. But Hume’s experiments take us in a step by step fashion from inanimate matter to sentient persons. This leaves his opponents with the burden of explaining the relevant differences between our everyday ascriptions of persistence across these domains.

This challenge has special force against soul theorists, such as Butler and Reid, who maintain that judgments of continued existence for persons are made on the basis of perfect or “strict” identity, whereas ascriptions of persistence for animals or vegetables are founded on imperfect identity and a “loose” and popular way of speaking (Butler, 2008, pp. 100–101; cf. Reid, 2008, pp. 111–112). Hume’s experiments reveal that one cannot draw

a categorical distinction along these lines. These persistence judgments are based on similar kinds of perceptual sequences, involve similar principles of the mind, and generate similar types of attributions. If talk of statically persistence in the case of biological organisms must be interpreted in a loose and popular sense, then we must adopt an equivalent stance with regard to the continued existence of selves.

3. Persistent objects: A cognitive science perspective

Hume approaches questions about personal identity as a scientist of human nature. We cannot properly understand what we ordinarily believe, or why we believe what we do, from the comforts of the philosophical armchair. Rather, we must carefully observe the types of attributions regularly made in common life and test rival hypotheses that purport to explain these facts.

There are serious concerns, however, about Hume's experimental methodology. Hume maintains that "we" are naturally inclined to make attributions of static persistence when presented with the four change vignettes (T 1.4.6.7–1.4.6.12; SBN, pp. 254–257). But these assertions, by our current lights, are too speculative. In particular, Hume provides no empirical evidence for the claim that our responses to these vignettes would be uniform in nature. He writes at one point that his proposals are based on "daily experience and observation (T 1.4.6.7; SBN, p. 255). But we are asked to take his word on this matter: he does not provide any records or data.

How would Hume's account fare if tested in studies with large and diverse samples of respondents? Fortunately, cognitive scientists have devised novel techniques for rigorously studying ordinary beliefs about persistence through change. This research allows us to evaluate the plausibility of Hume's proposals about the principles of mind that underwrite these commitments.

Consider Geoffrey Hall's research, for example, on folk attributions of persistence. Hall presents participants in his experiments with sequences of photographs depicting novel objects – described either as artifacts or organisms – whose parts are successively replaced over time. What he discovers is that respondents are significantly more likely to judge that these objects persist through change, even when there is a wholesale replacement of their parts, as long as these transformations occur *gradually* (Hall, 1998). It appears that Hume was on the right track in proposing, therefore, that our everyday attributions of persistence to artifacts and biological organisms does not require that they contain any of their original parts; it suffices that the new parts are introduced in an incremental fashion.

David Rose's recent work in experimental philosophy also provides evidence in favor of Hume's proposal concerning the conditions under which we ordinarily attribute persistence to objects that suddenly change.

Rose presents participants with a series of vignettes about material objects that undergo instantaneous alterations in their functional properties. Half of the participants are informed that the purpose of these objects is preserved; the other half are told that their function is transformed. Rose discovers that individuals in the ‘preserved function’ condition are significantly more likely to judge that the same object persists over time (Rose, 2015).⁶ Hume’s conjecture has once again been corroborated: objects are believed to persist through sudden changes as long as they continue to perform the same functions.

Recent work in cognitive science also provides crucial support for Hume’s claim that everyday beliefs about diachronic persistence depend on psychological biases rather than rational insight or evidence. Hall maintains that attributions of persistence to objects composed of entirely new parts is a by-product of our natural tendency toward *psychological essentialism*. We ordinarily think of whole objects as being greater than the sum of their parts, as he puts it, because we are inclined to believe that they have hidden natures or essences which make them the individual things they are (Hall, 1998). Children as young as age four “naturally assume” that objects contain essential cores that remain fixed and invariant over time (Medin & Ortony, 1989). This psychological “bias” induces us to think of these objects as having a “true nature” that remains “unchanged” even as their perceived characteristics are radically transformed (Gelman, 2004).

Psychological essentialism does not only underwrite our everyday beliefs about natural kinds, then, but it also manifests itself in how we think about the persistence of individual entities. Young children typically believe that the individual essences of animals remain fixed and invariant, for example, even when their observed behaviors and surface appearances undergo radical transformations (Gutheil & Rosengren, 1996). And mature adults display similar biases. Individual artifacts are frequently viewed as having underlying essences that constitute their identities and which make them preferable to perceptually indistinguishable duplicates (Hood, 2014; cf. Hood & Bloom, 2008). It is not only philosophers, therefore, who posit hidden and immutable cores. Human beings appear to have a natural propensity to suppose, without any evidence, that there are fixed essences that lie behind the manifold changes we perceive.

4. Persistent selves: A cognitive science perspective

Hume adopts an indirect strategy in defense of his hypothesis concerning the psychological mechanisms that are responsible for everyday beliefs about persistent selves. He takes himself to have securely established that attributions of persistence to vegetables and animals depends on biases and propensities of the imagination. We do not need to directly investigate

judgments of personal persistence, then, since we can make inferences on the basis of the “great analogy” between biological organisms and selves (T 1.4.6.5; SBN, p. 253).

The identity, which we ascribe to the mind of man, is only a fictitious one, and of a like kind with that which we ascribe to vegetables and animal bodies. It cannot, therefore, have a different origin, but must proceed from a like operation of the imagination upon like objects. (T 1.4.6.15, SBN, p. 259).

Hume assumes that there are no relevant differences in our beliefs about diachronic persistence across categories or domains. Once it has been demonstrated that our common attitudes about the persistence of plants and animals depends on psychological biases rather than rational insight, therefore, it can be safely concluded that the same is probably true of our everyday thinking about personal identity over time.

Recent work in cognitive science makes it clear that we can go one step further, however, and directly probe our lay intuitions about the persistence of persons. Consider the innovative studies, for example, designed by Newman et al. (2014). These researchers present participants in the United States with a series of cases about individuals who undergo radical changes in their beliefs and behaviors. Consider their vignette describing a father who is suddenly transformed into a responsible parent:

Al used to be a ‘deadbeat’ dad. In the past, he never showed any real affection for his children and never expressed any interest in his children’s lives. Now, however, Al is a very caring and involved father. (Newman et al., 2014)

Participants are subsequently presented with a forced choice task in which they are asked to decide whether these changes are caused by Al’s “true self” or “surface self”; they are also prompted to rate the extent to which this dramatic conversion is “true to the deepest, most essential aspects of [Al’s] being” (ibid.: p. 204).

What these researchers discover is that participants have a general tendency to interpret Al’s transition to being a caring father in terms of the emergence of a “deeper, hidden element” in his personhood (ibid.). Al did not, in other words, really change. He was always a devoted father; he just did not act like one. Participants are inclined to believe that Al now acts in accordance with an essential or true self that was “always hidden deep inside” (Strohming et al., 2017).

Folk judgments concerning personal persistence display another hallmark of essentialist reasoning: true selves are typically regarded as being *immutable*. When participants in experimental studies are queried about what they imagine their lives to be like in thirty years, for example, they have a strong tendency to report that their core selves will remain exactly the same (Christy et al., 2019). The same applies when participants are asked what their lives

would be like if, counterfactually, they had been born in a different historical period: the experimental results indicate that people expect their true or essential selves to remain fixed and invariant across different contexts.

Additional evidence in support of the proposal that diachronic identity is commonly understood in terms of immutable essences comes from research on the attitudes of the elderly toward persistence across lifetimes. Troll and Skaff (1997) conduct structured interviews with individuals (N = 150) aged 85 or older. Members of this age group are most likely to have undergone major changes and upheavals, such as the loss of loved ones, hospitalization, or a sudden change of residence. But when those in this population are asked whether they have remained the same over the course of their whole lives, nearly three quarters of them respond that their core or true self has remained fixed. Even though many of these respondents believe that their personality traits and behaviors have varied over time, they did not interpret these alterations as providing any indication that their deeper or essential self has changed. As one 103-year old woman in their study put it: “My core has stayed the same. Everything else has changed.” (Troll & Skaff, 1997)

Cognitive science research provides *prima facie* support, then, for Hume’s claim that everyday belief in statically persistent selves is based on psychological biases rather than rational insight or evidence. Human beings have a natural tendency to understand the persistence of entities – whether they are material objects, artifacts, biological organisms, or persons – in terms of essential cores that somehow resist change. Indeed, contemporary researchers take their studies to indicate that our everyday beliefs about the self are based on *fictions*.

[T]he true self is, shall we say, evidence-insensitive . . . Yet people have little trouble imbuing it with a host of hidden properties . . . The true self is posited rather than observed. It is a hopeful phantasm . . . The notion that there are especially authentic parts of the self, and that these parts can remain cloaked from view indefinitely, borders on the superstitious. This is not to say that lay belief in a true self is dysfunctional . . . Perhaps it is a useful fiction . . . But, in our view, it is a fiction nonetheless. (Strohming et al., 2017)

Belief in true or essential selves is fictional in the sense that it goes beyond the evidence. We have no reason or justification for supposing that there are unperceived cores deep within ourselves that remains fixed and immutable over time; we only arrive at these commitments because of a natural psychological tendency to categorize the world in terms of hidden essences.

5. Persistent selves: Cross-cultural diversity

If everyday beliefs about personal persistence are based on domain-general principles such as psychological essentialism, one would expect these ontological commitments to be uniform around the globe. Julian De

Freitas and colleagues (De Freitas et al., 2017) investigate the question of whether belief in true selves is universal by presenting change vignettes – including the case of Al who is suddenly transformed into a caring and involved father – to members of four different countries: one with a highly independent social orientation (United States) and three with more interdependent orientations (Singapore, Colombia, and Russia). The study design remains exactly the same: participants are asked to choose whether the personal changes described in the vignettes are caused by true or superficial selves, and they are asked to rate the extent to which these alterations conform to the “deepest, most essential aspects” of these persons (De Freitas et al., 2017).

Their experimental results point to a “noticeable cultural difference” in the degree to which members from these different groups adopt an essentialist stance toward the self (ibid.: p. 18). Populations from Colombia and Singapore are forty percent less likely than populations in the United States, for example, to interpret the change vignettes in terms of hidden or true selves (ibid.: p. 17). These findings are consistent with other recent cross-cultural studies. Researchers have discovered that Japanese populations (with a more interdependent social orientation), for example, are significantly less likely to believe in essential core selves than Australian populations (with a more independent social orientation) (Tsukamoto et al., 2015). Members of cultures with interdependent social orientations do not, in general, exhibit a strong tendency to construe the boundaries of selves as stable and fixed independently of context (cf. Kung et al., 2016; Varnum et al., 2010). Even if participants from the United States are disposed to postulate essential core selves, then, populations with different cultural backgrounds do not reliably share these intuitions.

Similar cognitive disagreements have been found when comparing American and Tibetan populations. Shaun Nichols and his colleagues present a large sample [N = 520] of lay persons from the United States and Tibet with survey questionnaires that probe their commitment to immutable core selves. They ask participants to rate the degree to which they agree (on a 1–7 Likert scale) with the following two statements about their “core self beliefs”:

[1] *There is some essential core that is your self that stays exactly the same throughout your life.*

[2] *My core self will be exactly the same in a year as it is now.* (Nichols et al., 2018)

The results of these surveys reveal significant disparities between these populations: lay Tibetans tend to reject the notion of core persisting selves, whereas their American counterparts on average put credence in it (ibid.: p. 7).

Michael Chandler and his colleagues provide further evidence for cultural diversity in beliefs about the continued existence of selves. These researchers conduct “personal persistence” interviews with participants from both First Nations (Indigenous) and culturally mainstream populations in Canada. They first present these informants with a series of fictional stories in which protagonists undergo major changes, such as the transformations undergone by Scrooge in *A Christmas Carol* or those of Jean Valjean in *Les Misérables*, and subsequently ask whether they believe that these characters remain one and the same from the beginning to the end of these stories and why they believe this is the case. They also ask participants about any personal changes they have experienced and investigate their justifications for believing in their own personal persistence, or lack thereof, across the duration of their lives.

What these researchers find is that a significant majority of those in the cultural mainstream interpret the persistence of these fictional characters, as well as their own selves, through the lens of an “unchanging entity or essential core” (Chandler & Proulx, 2008). There is a widespread tendency in this population to suppose, as these authors put it, that a “subterranean . . . unchanging core” remains static beneath these perceived changes (Chandler et al., 2003).⁷ But this essentialist stance is significantly less prevalent among First Nations youth; indeed, less than 20% of this population attribute these personal changes to the influence of hidden or immutable selves (Chandler & Proulx, 2008).⁸ Commitment to essential core selves is not global, then, but rather is shaped by local cultural factors.⁹

6. Narrative selves and useful fictions

First Nations respondents do in fact typically believe that they remain numerically the same over long stretches of time. The crucial point is that they do not account for their continued existence by invoking any immutable pearls or hidden self-kernels. Rather, they have a strong tendency to justify their belief in personal persistence in terms of whole life stories. A majority (80%+) of the First Nations participants – and a minority (15–20%) of respondents from the cultural mainstream – maintain their sense of personal continuity via a “bridging narrative” that assimilates the various changes they experience into the structure and format of life stories (Chandler & Proulx, 2008; cf. Chandler et al., 2003).¹⁰

Research in cognitive science suggests that it is not unusual for people to view themselves as protagonists in ongoing life stories (D. P. McAdams & Adler, 2010; McAdams, 2006). If you ask a three-year old about what happened during the recent past, you are likely to receive a sequential list of remembered events. But the ability to represent important episodes of our lives in narrative format typically emerges during adolescence (Habermas &

Bluck, 2000). The rough outlines of these whole life stories are continually revised and edited, and eventually accepted or rejected, during our later adult years (McAdams, 2006).¹¹

When we think of our lives in terms of whole life stories, the personal past is no longer portrayed as serial chronicles or historical annals. Conscious experience is successive; it is just one event after another. But narrative selves are not represented this way. Rather, plot lines are introduced; obstacles and challenges come into focus; a select group of characters are cast as heroes or villains; key past scenes are rendered as turning points (McAdams, 1996, pp. 311–312). Life stories are typically organized, moreover, around general themes. Two themes are especially prevalent in the narratives articulated by adults in the United States. One prototypical theme involves “redemption sequences”, where painful events in the past are compensated by their contribution to an emotionally positive present; another recurrent motif involves “contamination sequences”, where the narrative arc bends in the opposite direction: pleasant scenes from the past are spoiled by current pain and suffering (McAdams & McLean, 2013).

Life stories serve to smooth over, or buffer, sudden biographical changes. When adolescents are queried, in interview settings, about disruptions in their lives, such as abrupt changes in personal relationships, they often feign or postulate narrative connections in order to help make sense of these transitions (Habermas & De Silveria, 2008). Indeed, one can observe a direct correlation between the quantity of perceived change and the number of narrative links woven into life stories (Habermas & Kober, 2015, p. 670). The story format provides structure to event sequences that are otherwise experienced as fragmentary or disjointed. As one leading researcher in the field of narrative psychology puts it, whole life stories help to construct a “patterned life out of what may appear, at first blush, to be a random and scattered life” (McAdams, 1996, p. 309).

Life stories are fictional in the sense that they are created in order to supply our lives with a greater sense of unity and coherence. This is not to say that people ordinarily regard these narratives as fictitious; they genuinely believe the stories they tell. And these life stories, except in exotic cases, are not pure fantasies. They are usually based on matters of fact whose veracity can be corroborated or contradicted by others (Bruner, 2004, p. 693). Nonetheless, these narratives are not objective life histories. And there are reasons to be suspicious about their reliability. One consideration lies with recent findings about the accuracy of autobiographical memories. Episodic memories are not videotapes that accurately reproduce previous events or states of affairs; rather, these memories often reconstruct the past in light of our present concerns and future goals (D. P. McAdams & Adler, 2010, p. 40). People are often highly selective, moreover, in choosing the characters, scenes, or themes that organize their life stories (Walker, 2012,

pp. 64–5). Finally, these narratives are often motivated: individuals with greater affective concern for the well-being of others, for example, are more likely to formulate their life stories according to themes of redemption rather than contamination (McAdams & Guo, 2015, p. 479).

To the extent that narrative selves are fictitious, however, they appear to be *useful fictions*. Researchers have observed a strong correlation between life stories and psychological well-being. But it matters what type of story one tells. Individuals who formulate their narratives around themes of redemption, for example, are more likely to report higher levels of happiness; going in the opposite direction, those who articulate dramatic changes in terms of contamination sequences are more likely to be depressed (cf. Adler et al., 2015; McAdams & Guo, 2015). Canadian adolescents who portray major turning points of their lives as having positive resolutions, moreover, report significantly higher levels of psychological well-being than those who do not (Tavernier & Willoughby, 2012). Finally, patients in psychotherapy who construe personal changes in terms of themes of agency – where they portray themselves as protagonists battling obstacles – tend to exhibit better mental health (Adler, 2012).

Belief in true or essential selves, it should be pointed out, also appears to be a convenient fiction. These commitments go beyond the available evidence, but they have positive influences on subjective well-being. Recent empirical studies based on large samples of American populations [N = 299] have established, for example, that those who posit immutable hidden selves typically report higher levels of life-satisfaction than those who do not (Dulaney et al., 2019; De Freitas et al., 2017). Essentialist and narrative tendencies should be seen as alternative psychological strategies, therefore, for resisting or defying personal change and providing individuals with a greater sense of unity and coherence in their lives. The essentialist strategy accomplishes this feat by positing a hidden substantial entity that remains immutable in an ontological realm independent of our successive perceptions. The narrative strategy performs this feat by accommodating personal changes, and connecting the diverse chapters of our lives, according to the representational format of whole life stories.

Failure to adopt either of these psychological heuristics can be hazardous. Consider once again the research on First Nations adolescents conducted by Chandler and colleagues. Members of this population typically interpret persistent selves in narrative terms. But whole life stories are harder to articulate for Indigenous groups whose cultural traditions and oral histories have been destroyed by colonization. The prevalence of suicide in these groups is disproportionately high. And this tragic situation appears to depend, to some extent, on their inability to tell coherent life stories. Eighty-percent of young adults in these high suicide-risk populations lack any psychological strategies for accounting for their personal persistence

(Chandler & Proulx, 2008; cf. Chandler & Ball, 1990). These adolescents are no longer able to accommodate disruptions or imagine their futures; they suffer, as these researchers put it, a “loss of temporal connectedness” (Chandler & Proulx, 2008). Unable to tell their whole life stories, it seems, they are literally falling apart.

Research in cognitive science vindicates Galen Strawson’s descriptive claim that there are individual “variations . . . of temporal temperament” or “time-style[s]” (Strawson, 2004; cf., 2017, pp. 109–110). People do not universally experience their lives in story-like terms. But recent work also casts doubt on Strawson’s normative proposal that “episodic” individuals – who do not regard themselves as having a continued existence over long periods of time – are often just as happy as “diachronics” who see themselves as persistent persons (ibid: pp. 432–3). Empirical research suggests that the belief in personal persistence across our lives substantially increases psychological well-being. This is not to deny that Strawson or others are happy episodics; it is just to say that they might be lucky exceptions.

7. Conclusion: Natural belief in selves

According to Terence Penelhum’s influential interpretation, Hume’s account of personal identity is a muddle. Hume assumes that qualitative changes are incompatible with numerical identity over time. But this is an “elementary error” in conceptual analysis (Penelhum, 1955, p. 577). Just as a single tune is composed of a succession of different notes, so one person consists of a series of alternating perceptions (ibid: p. 580). Personal persistence does not, in other words, require static invariance. Hume’s skepticism about belief in persistent selves turns on an overly restrictive account of the conditions of diachronic identity.

Penelhum’s critique, however, misses its mark. Hume approaches questions of personal identity, as we have seen, as a scientist of human nature. His ambition is not to introduce or defend a standard for re-identifying persons over time. And he does not endorse the narrow criterion that selves are temporarily extended if and only if they remain completely frozen in time. It is the untutored layperson, according to Hume, who typically believes in statically persistent selves. Hume’s experimental project is to understand the principles of mind that are responsible for this widespread commitment.

Hume’s main proposal is that there is a close resemblance between how we ordinarily think about the persistence of selves and biological organisms. How well do this hypotheses fare when evaluated from a cognitive science perspective? To the extent that folk beliefs about persistent selves and biological organisms are driven by domain-general principles such as psychological essentialism, Hume’s analogy appears quite strong. But to the

extent that these ordinary beliefs are products of a propensity to construe our lives in terms of whole life stories, this analogy appears weak. Beliefs about the persistence of biological organisms and selves, according to narrative psychologists, depend on fundamentally different types of mental operations. We ordinarily regard the nature of external objects in a ‘paradigmatic’ mode of thought, as Jerome Bruner puts it, whereas we typically regard ourselves in a ‘narrative’ mode (Bruner, 1986, pp. 12–13).

Cognitive science research on essential and narrative selves provide a good deal of support for the Humean proposal that everyday attributions of personal persistence are based on biases and propensities of the imagination. The human mind has a natural tendency, as far as we can tell on the basis of experiments, to *disguise* the perpetual flux of conscious experience. This imaginative feat is accomplished by either postulating hidden and immutable entities that defy change (essentialist strategy) or it assimilating these changes into the format of whole life stories (narrative strategy). The gap between what we perceive and believe about ourselves is filled by fictional elements. It is not rational insight or evidence, therefore, that underlies our ordinary sense of unity over time.

It is not surprising that commitment to persistent selves, therefore, is extremely difficult – if not impossible – to extirpate through explicit reasoning. Consider the research of Shaun Nichols and colleagues, once again, on Buddhists attitudes toward the diachronic persistence of selves. Monastic Buddhist monks are officially committed to a ‘no-self’ view, according to which selves are transient or punctual: they do not continue to exist from moment to moment, let alone across entire lifespans (Nichols et al., 2018, p. 9). But when these monks are queried about their attitudes toward death, they typically report feeling intense fear of personal annihilation (ibid: p. 10). Commitment to the persistence of selves appears “resistant” to philosophical arguments or spiritual exercises (ibid: p. 16). Just as the most ardent skeptic about induction cannot stop believing in floors as they step out of their beds, so too the most devout Buddhist monk or no-self theorist view cannot help but view themselves as persistent beings when they quit their meditations and return to everyday life.¹² These are *natural beliefs*, in Humean terms, that come to us involuntarily, like breathing.

Notes

1. In what follows, I will use the following abbreviations for Hume’s works: *A treatise of human nature* (Hume, 2002) as T; and the Selby-Bigge edition of *A treatise of human nature* (Hume, 1978) as SBN.
2. We ordinarily believe that there are many aspects of the self that undergo transformation over time. Students entering university do not expect to leave, for example, with exactly the same set of beliefs and desires. What Hume is claiming is that there we

commonly believe there is an essential core to the self which stands apart from these perceived changes. These immutable pearls or kernels are supposed to remain frozen in time in much the same way that the *60 minutes* logo spans the successive frames of the shows' episodes.

3. Associative principles play a productive role in this conflation. We do not observe anything in the contents of consciousness that remains unchanged over time; but we do perceive close relations between these fluctuating perceptions. Some of these relations are *causal*: perceptions “mutually produce, destroy, influence, and modify each other” (T 1.4.6.19; SBN p. 261). Some of these relations are based on *resemblance*: current memories and past perceptions are similar in terms of their representational content, after all, and differ only in their phenomenal force and vivacity. It is the multiple associative connections between the various items in our perceptual sequences that induces the mind to move easily over these changing sequences (AABBCCDDEE) and mistakenly view them *as if* they are with constant and invariable sequences (AAAAAAAAA). These are the “uniting principles”, as Hume puts it of the mental world (T 1.4.6.16; SBN, p. 260).
4. This sleight of hand resembles the ventriloquist trick where it appears to the audience as though the words are coming straight out of the mouth of the dummy. We believe that we see something, in other words, which is not really perceived. It is helpful to compare this sleight of hand to the causal illusion produced by the imagination. We commonly suppose that we apprehend necessary connections between events, according to Hume, even though we are only immediately acquainted with constant conjunctions. The analogy with causal illusions is helpful in another way. Philosophers are aware that we are not perceptually acquainted with necessary connections. But it is psychologically difficult to shake this illusion of the imagination, and as a result, they inevitably search in vain for these connections in the object (T 1.4.3.9; SBN, p. 223).
5. It is not only the “vulgar” or folk who allow themselves to be governed by such “trivial principles of the imagination”; even an illustrious thinker such as Lord Shaftesbury, according to Hume, is vulnerable to this psychological bias (T 1.4.6.6.; SBN, p. 255). Shaftesbury recognizes that the material parts of biological organisms are continually replaced over time; in order to justify the belief that particular plants or animals statically persist through this succession of parts, therefore, he supposes that there is some underlying and hidden “uniting principle” that remains fixed and immutable beneath these perceived changes (T 1.4.6.6 n; SBN, p. 50).
6. The Rose experiments reveal that we naively think that the teleological ends of objects are intrinsic to their natures. We are naturally “promiscuous teleologists”, as Kelemen and Rosset put it, with a natural tendency to interpret natural phenomena in terms of purposes (Kelemen & Rosset, 2009, p. 138). When this psychological tendency gets combined with psychological essentialism, it leads us to think about persistence of objects in terms of preserving functions.
7. Some culturally mainstream respondents regarded these transformations as a kind of metamorphosis, where a person gradually develops into their true or essential self in much the same way that a caterpillar realizes its potentiality while becoming a butterfly; others in this group thought of themselves in preformationist terms, where personal changes are understood in terms of the unfolding of a latent true self that was “already present.. from the very beginning” (Chandler et al., 2003, p. 33).
8. The Chandler and Proulx research demonstrates that the First Nations populations do not accommodate personal changes in terms of essential or true selves. This provides *prima facie* reason to deny they are ontologically committed to the existence of subterranean hidden selves. But it does not demonstratively prove that this is so. It

remains an open possibility that members of these Indigenous groups remain committed to hidden and immutable selves, even though they do not invoke these entities in their explanations of personal changes. I would like to thank an anonymous reviewer for bringing this point to my attention.

9. One might wonder how Hume's claim about the *natural* foundations of belief in persistent persons can be reconciled with the claim that *cultural* factors play a crucial role in determining whether individuals or groups come to view themselves in narrativist or essentialist terms. Hume's official position is that human beings have a *general* propensity to believe that they persist over stretches of time; but cognitive research indicates that local conditions on the ground help determine the *particular* conceptions of persistence that emerge across different societies. It should be noted that this position closely resembles Hume's account of the natural foundations of morality, according to which human beings universally share the same set of *fundamental* principles, but different beliefs and socio-economic conditions generate *superficial* diversity of moral codes across societies.
10. Some members of the First Nations population leaned toward episodic or picaresque type of narratives; these stories involve minimal connections; what holds their lives together is nothing but the "passing of time" (Chandler et al., 2003, p. 39). Other respondents adopted a fuller narrative account where they see themselves as authoring or interpreting their lives in terms of chapters in an ongoing story (ibid: pp. 40–42).
11. It should be noted that Lorenzo Greco and Tony Pitson offer narrativist interpretations of Hume's theory of personal identity in Books Two and Three of the *Treatise*. Greco maintains that the stories we tell ourselves about our moral agency play a role, according to Hume, in helping us see ourselves as a "cohesive whole" rather than a merely disconnected bundle (Greco, 2015, p. 708). It is not clear whether Greco is claiming that Hume *in fact* offers a narrativist account of personal persistence, or if he is claiming that this is what Hume *should* have said. Greco writes at one point, for example, that Hume's account of the self "bears some strong similarities" to a narrative conception of selves (Greco, 2015, p. 700); elsewhere, he refers to the narrative account as "Humean" in spirit (ibid.: p. 707, cf. p. 711). Pitson is more explicit that he is "pursuing this claim on [Hume's] behalf" (Pitson, 2002, p. 93).
12. Tibetan Buddhist Monks do not explicitly endorse an essentialist view of the self; they see persons, rather, as changing bundles of five aggregates. The fact that they fear personal annihilation entails that they are implicitly committed to their continued existence; but it does not entail, as Nichols and his colleagues suggest, that they adopt an essentialist attitude toward their core selves. There is an excluded alternative: a narrative approach to personal persistence. As Nichols and his colleagues point out, after all, autobiographies are a paradigmatic genre in Tibetan literature (Nichols et al., 2018, pp. 16–17). So the monastic Tibetans might make the supposition that persistent persons are composed of narratively connected bundles of perceptions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Mark Collier is Professor of Philosophy at the University of Minnesota, Morris and Affiliate Faculty Member at the University of Minnesota Center for the Cognitive Sciences. He previously served on the faculty at Stanford University, the University of Pittsburgh, Pomona College, and Peking University. His research focuses on examining Hume's experimental philosophy from a cognitive science perspective.

References

- Adler, J. M. (2012). Living into the story: Agency and coherence in a longitudinal study of narrative identity development and mental health over the course of psychotherapy. *Journal of Personality and Social Psychology*, 102(2), 367–389. <https://doi.org/10.1037/a0025289>
- Adler, J. M., Turner, A. F., Brookshier, K. M., Monahan, C., & Walder-Biesanz, I. (2015). Variation in narrative identity is associated with trajectories of mental health over several years. *Journal of Personality and Social Psychology*, 108(3), 476–496. <https://doi.org/10.1037/a0038601>
- Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press
- Bruner, J. (2004). Life as narrative. *Social Research*, 71(3), 691–710.
- Butler, J. (2008). Of personal identity. In J. Perry (Ed.), *Personal Identity* (pp. 99–105). University of California Press.
- Chandler, M. J., & Ball, L. (1990). Continuity and Commitment: A Developmental Analysis of the Identity Formation Process in Suicidal and Non-suicidal Youth. In H. A. Bosma & A. E. S. Jackson (Eds.), *Coping and Self-Concept in Adolescence* (pp. 149–166). Springer.
- Chandler, M. J., & Proulx, T. (2008). Personal Persistence and Persistent Peoples: Continuities in the Lives of Individual and. In W. C. Communities & F. Sani (Eds.), *Self-Continuity: Individual and Collective Perspectives* (pp. 213–226). Psychology Press.
- Chandler, M. J., Lalonde, C. E., Sokol, B., & Hallet, D. (2003). Personal persistence, identity development and suicide: A study of Native and non-Native North American adolescents. *Monographs of the Society for Research in Child Development*, 68(2), 1–128.
- Christy, A., Schlegel, R., & Cimpian, A. (2019). *Why do people believe in true selves? The role of psychological essentialism*. (Unpublished manuscript)
- De Freitas, J., Sarkissian, H., Newman, G. E., Grossman, I., De Brigard, F., Luco, A., & Knobe, J. (2017). Consistent belief in good true self in misanthropes and three independent cultures. *cognitive science*, 1–27.
- dulaney, e.s., graupmann, v., quinn, k.a. (2019). who am i and how often?: variation in self-essentialism beliefs, cognitive style, and well-being. *Personality and Individual Differences*, 136(1), 148–159 .
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Science*, 8(9), 404–409. <https://doi.org/10.1016/j.tics.2004.07.001>
- Greco, L. (2015). The self as narrative in Hume. *Journal of the History of Philosophy*, 53(4), 699–722. <https://doi.org/10.1353/hph.2015.0086>
- Gutheil, G., & Rosengren, K. S. (1996). A rose by any other name: Preschoolers understanding of individual identity across name and appearance changes. *British Journal of Developmental Psychology*, 14(4), 477–498. <https://doi.org/10.1111/j.2044-835X.1996.tb00719.x>

- Habermas, T., & Bluck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin*, 126(5), 748–769. <https://doi.org/10.1037/0033-2909.126.5.748>
- Habermas, T., & De Silveria, C. (2008). The development of global coherence in life narratives across adolescence: Temporal, causal, and thematic aspects. *Developmental Psychology*, 44(3), 707–721. <https://doi.org/10.1037/0012-1649.44.3.707>
- Habermas, T., & Kober, C. (2015). Autobiographical reasoning in life narratives buffers the effect of biographical disruptions on the sense of self-continuity. *Memory*, 23(5), 664–674. <https://doi.org/10.1080/09658211.2014.920885>
- Hall, D. G. (1998). Continuity and the persistence of objects: When the whole is greater than the sum of the parts. *Cognitive Psychology*, 37(1), 28–59. <https://doi.org/10.1006/cogp.1998.0688>
- Hood, B. (2014). Quiddity and haecceity as distinct forms of essentialism. *Behavioral and Brain Sciences*, 37(5), 492–493. <https://doi.org/10.1017/S0140525X13003762>
- Hood, B., & Bloom, P. (2008). Children prefer certain individuals over perfect duplicates. *Cognition*, 106(1), 455–462. <https://doi.org/10.1016/j.cognition.2007.01.012>
- Hume, D. (1978). *A treatise of human nature* (2nd ed.). Oxford: Clarendon
- Hume, D. (2002). *A treatise of human nature*. Oxford University Press.
- Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanations in adults. *Cognition*, 111(1), 138–143. <https://doi.org/10.1016/j.cognition.2009.01.001>
- Kung, F. Y. H., Eibach, R. P., & Grossmann, I. (2016). Culture, fixed-world beliefs, relationships, and perceptions of identity change. *Social Psychological and Personality Science*, 7(7), 631–639. <https://doi.org/10.1177/1948550616652208>
- McAdams. (2006). *The Redemptive Self: Stories Americans Live By*. Oxford University Press.
- McAdams, D. P. (1996). Personality, Modernity, and the Storied Self: A contemporary framework for studying persons. *Psychological Inquiry*, 7(4), 295–321. https://doi.org/10.1207/s15327965pli0704_1
- McAdams, D. P., & Adler, J. M. (2010). Autobiographical memory and the construction of a narrative identity: Theory, research, and clinical implications. In J. E. Maddux & J. P. Taguey (Eds.), *Social psychological foundations of clinical psychology* (pp. 36–50). Guilford Press.
- McAdams, D. P., & Guo, J. (2015). Narrating the generative life. *Psychological Science*, 26(4), 475–483. <https://doi.org/10.1177/0956797614568318>
- McAdams, D. P., & McLean, K. C. (2013). Narrative identity. *Current Directions in Psychological Science*, 22(3), 233–238. <https://doi.org/10.1177/0963721413475622>
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge University Press.
- Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality & Social Psychology Bulletin*, 40(2), 203–216. <https://doi.org/10.1177/0146167213508791>
- Nichols, S., Strohminger, N., Rai, A., & Garfield, J. (2018). Death and the Self. *Cognitive Science*, 42, 1–19. <https://doi.org/10.1111/cogs.12590>
- Penelhum, T. (1955). Hume on personal Identity. *Philosophical Review*, 64(4), 571–589. <https://doi.org/10.2307/2182635>
- Pitson, A. E. (2002). *Hume's Philosophy of the Self*. Routledge.
- Prinz, J. J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford University Press.
- Reid, T. (2008). Of identity. In J. Perry (Ed.), *Personal Identity* (pp. 107–112). University of California Press.

- Rose, D. (2015). Persistence through function preservation. *Synthese*, 192(1), 97–146. <https://doi.org/10.1007/s11229-014-0555-6>
- Strawson, G. (2004). *Against narrativity*, *Ratio (new series)*, 17, 428–452.
- Strawson, G. (2017). *The Subject of Experience*. Oxford-New York: Oxford University Press.
- Strohming, N., Knobe, J., & Newman, G. (2017). The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science*, 12(4), 551–560. <https://doi.org/10.1177/1745691616689495>
- Tavernier, R., & Willoughby, T. (2012). Adolescent turning points: The association between meaning-making and psychological well-being. *Developmental Psychology*, 48(4), 1058–1068. <https://doi.org/10.1037/a0026326>
- Troll, L. E., & Skaff, M. M. (1997). Perceived continuity of self in very old age. *Psychology and Aging*, 12(1), 162–169. <https://doi.org/10.1037/0882-7974.12.1.162>
- Tsukamoto, S., Holland, E., Haslam, N., Karasawa, M., & Kashima, Y. (2015). Cultural differences in perceived coherence of the self and ingroup: A Japan-Australia comparison. *Asian Journal of Social Psychology*, 18(1), 83–89. <https://doi.org/10.1111/ajsp.12090>
- Varnum, M. E. W., Grossman, I., Kitayama, S., & Nisbett, R. E. (2010). The origin of cultural differences in cognition: Evidence for the social orientation hypothesis. *Current Directions in Psychological Science*, 19(1), 9–13. <https://doi.org/10.1177/0963721409359301>
- Walker, M. J. (2012). Neuroscience, self-understanding, and narrative truth. *AJOB Neuroscience*, 3(4), 63–74. <https://doi.org/10.1080/21507740.2012.712603>