Sim Bamford
and John Danaher

Transfer of Personality to a Synthetic Human ('Mind Uploading') and the Social Construction of Identity

Abstract: *Humans have long wondered whether they can survive the death of their physical bodies. Some people now look to technology as a means by which this might occur, using terms such 'whole brain emulation', 'mind uploading', and 'substrate independent minds' to describe a set of hypothetical procedures for transferring or emulating the functioning of a human mind on a synthetic substrate. There has been much debate about the philosophical implications of such procedures for personal survival. Most participants to that debate assume that the continuation of identity is an objective fact that can be revealed by scientific enquiry or rational debate. We bring into this debate a perspective that has so far been neglected: that personal identities are in large part social constructs. Consequently, to enable a particular identity to survive the transference process, it is not sufficient to settle age-old philosophical questions about the nature of identity. It is also necessary to maintain certain networks of interaction between the synthetic person and its social environment, and sustain a collective belief in the persistence of identity. We defend this position by using the example of the Dalai Lama in Tibetan Buddhist tradition and identify technological procedures that could increase the credibility of personal continuity between biological and artificial substrates.*

Correspondence:
John Danaher, School of Law, NUI Galway, University Road, Galway, Ireland.
Email: john.danaher@nuigalway.ie

Journal of Consciousness Studies, **24**, No. 11–12, 2017, pp. ?–?

Keywords: personal identity; mind uploading; personality transfer; social construction; distributed cognition; transhumanism; singularitarianism.

1. Introduction

The terms ‘whole brain emulation’, ‘mind uploading’, and ‘substrate independent minds’ have been used in recent years to describe the hypothetical possibility of transferring or emulating the functioning of a human’s brain or ‘mind’ on a synthetic substrate. Philosophers and scientists debate both the technical feasibility and philosophical plausibility of such transfers. In this paper we take no view on the technical debate, except to note that the proposed methods are all highly speculative, relying on knowledge and techniques which have not been created.

The philosophical debate has largely been concerned with the question of identity: will a person survive the transferral process from a biological to a synthetic substrate? The answers to this question have settled into well-worn grooves, with participants presuming that there is some objective fact of the matter when it comes to the continuation of identity and that this can be ascertained through scientific or rational enquiry. In this paper, we argue that this understanding of the debate is flawed. What many participants miss is that personal identity is, in large part, a social construct. Whether the same identity survives a process of transfer depends on social factors. There is good reason to think that social factors will be more important than neurological or metaphysical ones when it comes to the future utilization of personality transfer technologies. We explain why this is the case, showing how our personal identities are constructed by social factors both causally and constitutively (terms which will be explained in Section 4), and how the choice of technical process for personality transfer may influence the continuation of personal identity.

The paper proceeds as follows. In Section 2, we briefly summarize some of the proposed procedures for mind uploading and brain emulation, suggesting that they are best described by the label of ‘personal transfer to synthetic human’. In Section 3, we highlight the importance of personal identity to the philosophical debates about these procedures, and introduce our argument about the social construction of identity. In Section 4, we clarify what we mean by social construction, distinguishing between two distinct forms. In Section 5, we defend the claim that personal identity is socially constructed, using an analogy

with the reincarnation of the Dalai Lama to make our point. Finally, in Section 6, we identify various features of personality transfer procedures that could make the continuation of identity more socially credible.

2. Proposed Procedures for Personality Transfer

In this section we briefly summarize the main possibilities that have been suggested in non-fictional writing that are linked to the concept of mind uploading. Some speculate (Moravec, 1988, p. 110; Kurzweil, 2000, pp. 52–4; Sandberg and Bostrom, 2008; Hayworth, 2010a; 2012; Koene, 2012a; Deca, 2012) that a human brain could be scanned at a high level of detail; that from these data its functioning could be derived and simulated by a computer, where the simulation may be at a neural level or a more abstract functional level; the simulation would then couple with a robotic or virtual body in order to reanimate the individual who was scanned (typically the scanning is envisaged for the brain of a deceased individual). Some focus (Koene, 2012b; Martins, Erlhagen and Freitas, 2012) on recording the low-level functioning of the live brain as a complementary data-gathering approach to assist in the parametrization of such simulations. Others speculate (Rothblatt, 2007; 2012; Bainbridge, 2009; 2012), that information collected about an individual (including but not limited to psychological self-analyses) could be used to parametrize a generic substrate, i.e. a humanoid robot, in order to reconstruct the individual. Yet others speculate that a human's brain (and body) could be gradually replaced by synthetic parts with the end result that the individual would continue to live with a synthetic (robotic) substrate (Moravec, 1988, pp. 109–10; Kurzweil, 2000, pp. 52–4); or that a device worn externally over an extended period could learn to emulate a particular person, so as to eventually control a robotic replacement for them (Moravec, 1988, p. 110). Wiley (2014) and Bamford (2012) have both attempted to taxonomize the different procedures and provide some analysis of the differences between them. It should be clear from their descriptions that each of these proposed procedures is highly speculative, relying on technology and know-how which has not been created.

All of these proposals involve some attempt to copy information about a person (the functionality of their brain; their personality traits, etc.) and reproduce this information in a synthetic substrate (robot,

digital computer, etc.). They vary in terms of the information they try to capture and reproduce, and the methodology they use to transfer this information from the original, biological person to the synthetic analogue. What they all have in common is that before the procedure there would be a live human (hereafter 'biological human'), after the procedure there would be some synthetic, cybernetic, or virtual human-like entity (hereafter 'synthetic human'), and there would be some reason to claim that the biological human and the synthetic human were the same person — the same identity.

That the continuation of personal identity is key to understanding the debate about these proposed procedures is clear from how they are described by the many authors who have discussed them. For example:

> The debate over mind uploading revolves around a central question, 'What do you consider to be you?' Mind uploading is useless if this personal definition of 'you' is not successfully transferred. (Hayworth, 2010a)

> Is there any chance that we — you or I, personally — can fully share in the magical world to come? This would call for a process that endows an individual with all the advantages of the machines, without loss of personal identity. (Moravec, 1988, p. 109)

> The… question… is whether the post-transfer computers will be the same persons as those who undertook the transfer. Will the identity of the person be maintained? (Phillips, 2000, chapter 5)

Rothblatt (2012) described an experiment which would allow us to decide if '…the software-based mind is a technoimmortalized continuation of the predecessor's identity'.

Oto (2012) identifies '…a vexing problem of personal identity — for we will wish to know if an uploaded copy of our mind amounts to a copy of us, or actually is us'.

Hopkins (2012) wrote:

> The real question in uploading is whether uploading procedures maintain the identity of the specific mind throughout the process.

The common theme in the above quotations is that the preservation of personal identity is the important factor. The terms 'mind uploading' and 'substrate independent minds' give the impression that the human mind is the target of these transfer procedures. But the term 'mind' can be ambiguous. In the sense used by the authors above it refers to a set of memories and patterns of thinking of an individual human that sets them apart from individuals in similar circumstances. Many of

these authors adopt a 'patternist' view of mind and personal identity (Wiley, 2014, chapter 5). They believe that identity is a question of preserving informational patterns. For example, Moravec wrote: 'Pattern identity… defines the essence of a person, say myself, as the pattern and the process going on in my head and body, not the machinery supporting that process' (1988, p. 117) (Goertzel, 2006, p. 2, traces patternist philosophy back further). Others express similar views:

> …mind is in essence about patterns of organization and behavior, and… the same patterns of organization and behavior can almost surely be realized via multiple different substrates. (Goertzel and Ikle, 2012)

> There is a pattern that is very dear to us. This pattern is the information content of our minds. By the information content, I mean both the parameter settings (e.g., memory), as well as the ways in which the parameters are used, the functions carried out by the mind… That pattern is all that we are aware of being. (Koene, 2011)

> …my identity is rather like the pattern that water makes when rushing around a rock in a stream. The pattern remains relatively unchanged for hours, even years, while the actual material constituting the pattern of the water is replaced in milliseconds. (Kurzweil, 2000)

So, according to these authors, one's 'mind' or 'identity' consists in a pattern of memories and mental functions that marks one out as a unique identity.

In order to choose an appropriate label under which to group the proposed procedures and the hope they represent, we prefer to set aside the labels 'mind uploading' and 'substrate independent minds', and also 'mind substrate transfer', which we previously endorsed (Bamford, 2012), and opt instead for 'personal transfer to a synthetic human (PTSH)'. This is a consequence of taking together the two points presented above in this section: (a) personal identity is considered the important target to be preserved through transfer; and (b) where 'mind' is considered the important target, there is a concurrent belief that 'mind' is a pattern which defines personal identity. We also favour PTSH over 'transmigration' (Moravec, 1988) and 'techno-immortalization' (Rothblatt, 2012) as it is more descriptive.

3. What is Personal Identity Anyway?

The debate about whether proposed procedures for PTSH would preserve personal identity is a branch of a broader and longer-running debate about the nature of the person. There are many competing

theories of personal identity, and no apparent consensus. Phillips (2000, chapter 5), for example, links the pattern-identity position of some of the aforementioned authors to the dominant class of Lockean theories of personal identity which include a memory criterion for identity (Locke, 1690/1975), and a psychological continuity criterion (Parfit, 1984, pp. 205–7). Schneider (2009, p. 7) points out that 'Patternism is an updated version of the psychological continuity theory'. Other authors have done a thorough job of both describing and highlighting alternative personal identity theories such as body identity, brain identity, and soul identity, according to which continuation of identity depends on maintaining the same body, brain, or soul. Animalist theories are also popular and maintain that the continuation of personal identity depends on the persistence of our animalistic properties (roughly: our homeostatic biological systems). If they cease to exist, we cease to exist (Olson, 2007). Biologically-oriented theories of this sort seem strongly opposed to the possibility of PTSH, hence why the Lockean, patternist theories are preferred among proponents of these technologies (Chalmers, 2014).

Irrespective of where these theories come down on the possibility of PTSH, they all assume that personal identity is constituted and sustained by the persistence of functionalist, psychological, biological, or supernatural factors (i.e. mental patterns, psychological states, biological processes, or a common 'soul'). There exists, however, an alternative view — that personal identity is a social construct. Kompridis (2009, p. 27) laid down the gauntlet:

> But what if personal identity is not in the head, not in the brain, and not something that can be extracted from the life history of an individual, rendered discrete, and subject to manipulable processes as is any mere 'object'? What if personal identity is constituted in, and sustained through, our relations with others, such that were we to erase our relations with our significant others we would also erase the conditions of our self-intelligibility?

In a complementary stance, Hongladarom (2011, p. 541) wrote:

> …personal identity has traditionally been associated with internalism — factors thought to be responsible for fixing the identity have come from internal sources such as the subject's own beliefs and memory episodes. However, one could follow the lead of the social epistemologists and other externalists in epistemology and… argue that external factors are really the ones that fix the identity… [O]ne could… try to locate the source instead outside of the subject's cognitive domain.

Such ideas are by no means widely accepted; indeed they go against commonly held notions that questions of continuity of personal identity are too important to be determined arbitrarily or to depend on third parties (Parfit, 1984, p. 267). But what we wish to argue in the remainder of this paper is that this scepticism is misplaced. Personal identity is indeed, in large part, a social construction. Whether we continue to exist depends not just on the satisfaction of psychological or biological conditions, it depends also on the satisfaction of social conditions, particularly conditions relating to networks of interaction between a human and their social peers, as well as systems of collective belief. What's more, when it comes to the practical utilization and significance of PTSH technologies, these social factors will have a far more decisive role to play than biological or psychological factors.

It helps in making this argument to distinguish between two perspectives on personal identity:

> **First-person perspective:** Do *I* continue to exist if I undergo a PTSH process? In other words, is the 'me' that starts the process the same as the 'me' that results from the process?
>
> **Third-person perspective:** Does *Jack* (or whoever) continue to exist if he undergoes a PTSH process? In other words, should we treat the biological *Jack* at the start of the process as being the same as the synthetic *Jack* at the end of the process?

The first-person perspective is concerned with the continuation of the self as a continuing subject of conscious experience. We are conscious and self-conscious beings. We are aware of what happens and we have a sense that it is happening to *us* (i.e. to some single, coherent self). We also sense that it is the same self that persists over time in having these experiences. When we look at PTSH from the first-person perspective what we are wondering is whether this self will survive the transferral process. The third-person perspective is different. It is concerned with the continuation of the self as a bundle of social roles, rights, duties, and responsibilities. We all stand in relations to one another. We are friends, lovers, bosses, enemies, and so forth. We have certain legal rights and duties towards others. We also stand in legally-recognized relationships to objects and persons in our social environments. When we ask the identity question from this perspective we are wondering whether we should ascribe the same bundle of social roles, rights, responsibilities, and duties to the synthetic

human that results from the transferral process. Whether we do, or not, obviously has enormous social significance. It will affect how we understand and accept putative PTSH technologies.

At first glance, it might seem like the senses of identity at play in both of these perspectives on the question are very different. Furthermore, it might be relatively easy to see how personal identity gets socially constructed from the third-person perspective, but less easy to see how it gets constructed from the first-person perspective. It is also probably fair to say that many of those interested in the debate about PTSH are concerned with identity from the first-person perspective. Nevertheless, we argue that a large part of personal identity is socially constructed no matter which perspective you take. This doesn't mean that biological or psychological factors are irrelevant, but their importance is conditioned by the social factors that construct identity.[1]

4. Two Varieties of Social Construction

Before we get into the meat of this argument, we need to clarify what is meant by social construction. Hacking (1999, p. 6) suggests that the core of any claim to social construction is a *contingency claim*. If you say that X is socially constructed, what you are saying is that X is not

[1] The authors of this paper differ, to some extent, in their commitment to the social constructionist thesis. Author 1 embraces a stronger version of the thesis in which social conditions are all that ultimately matter and that there exists no objective truth about the association of personal identities to particular humans but only beliefs about the personal identities of humans. Author 2 thinks that social conditions matter quite a lot — more than has been fully appreciated in this debate so far — but that the social constructionist thesis may still miss something important about personal identity. Consider, for example, the hypothetical perfect hermit, i.e. one that spends their entire life isolated from society, and not just one that retreats from society after being raised and socialized there. Would they still have an identity? Author 2 believes that such an individual would have some identity but that it is likely to be significantly impoverished and nothing like the kind of identity in which the proponent of PTSH is interested. Author 1 does not see a contradiction with the strong social constructionist thesis. The focus on social construction is simply the major implication of a fundamental view (discussed in Section 5, below) in which the set of concepts regarding a human which constitute the personal identity related to that human resides in the memories of humans. Since the set of humans whose memories can contain concepts about a particular human does not exclude that particular human whom the personal identity is a conceptualization of, that human would not lack an identity, but rather it would consist in a single conceptualization. In this case, there would be a lack of interactions in which the identity could develop, thus an identity which is impoverished in some ways with respect to the norm seems a likely result.

'determined by the nature of things' but rather by social factors. Diaz-Leon refines this to the claim that:

> **Social Construction:** If X is socially constructed, this means that the 'instantiation and distribution of X is contingent upon certain social events and arrangements: if those social events and arrangements were different, then facts about X could be different' (Diaz-Leon, 2013, pp. 2–3).

So when we claim that personal identity is socially constructed, what we are claiming is that the fact that any human body is associated with a particular identity (whether this is viewed from the first-person or third-person perspective) is contingent upon social events and arrangements. If those social events and arrangements had been different, the identity claim may not hold.[2]

We need to be even more precise. Most contemporary accounts of social construction distinguish between two types of social construction: the causal and the constitutive (Diaz-Leon, 2013; Haslanger, 2003; Mallon, 2008). The distinction is captured by the following:

> **Causal Social Constructionism:** X (some individual entity) is causally socially constructed as Y if and only if social factors play a significant role in causing X to have those features by which it counts as Y (adapted from Haslanger, 2003, p. 317, and Diaz-Leon, 2013, p. 5).

> **Constitutive Social Constructionism:** X (some individual entity) is constitutively socially constructed as Y if and only if X is a kind or sort Y such that in defining what it is to be Y we must make reference to social factors or, to put it another way, X is constitutively constructed as Y if social factors are metaphysically/conceptually necessary in our explanation or understanding of what it is for X to be Y (adapted from Haslanger, 2003, p. 318; Mallon, 2008, p. 6, and Diaz-Leon, 2013, p. 5).

---

2 Hacking (1999, p. 6) suggests that most social constructionist claims bring with them the implication that the current instantiation and distribution of X is bad or socially problematic, and so ought to be reformed, but these claims are not essential to the social constructionist thesis. It is possible to be a social constructionist about X and be perfectly happy about the instantiation and distribution of X. That is important to bear in mind here. When we claim that personal identity is socially constructed we are not making an evaluative claim about the desirability or triviality of this social arrangement. All we are saying is that identity is affected by social events and arrangements.

These definitions are best understood by way of examples. Any human-made artefact would count as *causally* socially constructed. Take a wristwatch (Diaz-Leon, 2013, p. 6). A wristwatch is a particular individual object and it belongs to the general class of wristwatches (devices that tell the time on your wrist). Clearly, the wristwatch didn't come into existence through spontaneous creation or a sequence of natural, non-human events. It was designed and fashioned by human beings, operating in particular social circumstances, at particular historical moments. It is, thus, causally socially constructed. Social factors have played a significant role in causing the wristwatch to have the features by which it counts as a wristwatch. But once the wristwatch has come into existence, it has some metaphysical independence from the social factors that were essential for its creation. The society that created it could crumble and fall and it would still exist as a wristwatch.

Contrast that with a case of constitutive construction. Social roles are the classic example. Take the status of being Prime Minister of the United Kingdom as an example. This status is determined entirely by social factors and events. It requires compliance with some formally agreed upon procedure (nomination, voting, election) and collective belief in and acceptance of the validity of that procedure. These factors *constitute* what it means to be Prime Minister. Without these social factors, a particular individual cannot be Prime Minister. These social factors sustain this status on an ongoing basis. If there is a political revolution or a change in the agreed upon procedure or process, the individual who used to be called Prime Minister will lose this status. This makes it quite unlike the case of the wristwatch. Being Prime Minister does not have conceptual or metaphysical independence from social factors. Those social factors are conceptually and metaphysically necessary if some individual is to count as Prime Minister.

These distinctions mean that there are three ways to interpret our claim that personal identity is socially constructed. It might mean: (a) that we think that in order for some human body to count as having a particular identity, certain social factors had to *cause* it to have the properties that we associate with that identity; or (b) it might mean that we think certain social factors actually *constitute* (sustain) the properties that enable it to count as having a continued identity; or (c) it might mean that we think identity claims involve both causal and constitutive construction. Our view is the latter. We think that in order for any individual body to count as having a continuing identity it

must be caused to have certain properties by social factors and must be sustained in having other properties by social factors. Roughly speaking, then, personal identity from a first-person perspective is primarily causally socially constructed (though not entirely, as we shall see), whereas personal identity from a third-person perspective is primarily constitutively socially constructed (where collective belief is the main social factor at play).

It is important to note that our position is not intended to be ethical or political in nature. As Hacking notes (1999, p. 6), claims to the effect that 'X is socially constructed' are often politically and ethically loaded. The person who makes that claim is usually appealing to some social injustice and asking for our practices of social construction to be reformed. We avoid all such ethical appeals here. We are arguing that identity is, in large part, socially constructed: we are not arguing that this is a good or bad thing. We do try to identify factors that are likely to contribute to the success or failure of an identity claim in the case of PTSH (Section 6, below), and we stand by our arguments in relation to those factors, but again our aim in doing this is not to defend the ethical propriety of those factors; it is to suggest that they are likely to work.

Our attempt at ethical neutrality is also complicated by the fact that identity claims are particularly politically loaded, with claims of the sort 'I am a man/woman/transgender, etc.' being among the most politically charged one can make. Part of the contention seems to be that Western societies are shifting toward a new norm whereby individual claims to identity are treated as being sacrosanct: if you say that you are X (man/woman, etc.), the society around you should respect that claim. This might be thought to create problems for our argument in so far as we suggest that identity claims are dependent on third-party/social factors.[3] But, again, we are not saying that this is a good thing, merely that it is a thing. And we would argue that whether or not we do successfully shift to a norm whereby identity claims are treated as sacrosanct is ultimately going to depend on the kinds of social factors we discuss below. So the shift to this norm would prove the very point we are trying to make.

There is a danger that our argument is perceived to be quite trivial. After all, in some sense it is trivially true that social factors are causally responsible for our identities: we would not be who we claim

[3] We are indebted to an anonymous reviewer for raising this point with us.

to be were we not caused to have certain psychological features by the societies in which we live. Likewise, it seems to be trivially true that social factors constitute our identities: the social roles, relationships, rights, and duties we have are sustained by various networks of collective belief. If you are a married property-owner — if those qualities are central to your identity — then you only have them to the extent that there is collective belief in certain legal rules and principles. If that collective belief is eroded, you would no longer have those qualities. What we want to argue next, however, is that there is nothing trivial about the type of social construction going on in both instances. They are, rather, core aspects of what it means for personal identity to persist over time.

5. How Personal Identity Gets Socially Constructed

The case of the Dalai Lama provides a useful analogy for our purposes. There is a belief in reincarnation in Buddhist culture. At its heart, this belief maintains that the same identity can be shared across different physical biological bodies. The most socially and politically significant instance of reincarnation comes in the shape of the Dalai Lama, the spiritual and political leader of the Tibetan people. The current Dalai Lama is the fourteenth in the line of succession. He was born Lhama Thondup on the 6 July 1935. Two years later he was recognized as the reincarnation of the thirteenth Dalai Lama and two years later again he was officially declared to be the fourteenth Dalai Lama.

Under Tibetan tradition, the Dalai Lama is a *tulku*, i.e. a human body that keeps or sustains a certain lineage-identity.[4] In the case of the Dalai Lama, all human bodies who are recognized as such are deemed to be incarnations of the Bodhissatva of Compassion. Officially, the line of Dalai Lama *tulkus* can be traced back to the fifteenth-century monk Gedun Drub, who was the first of the modern Dalai Lamas; unofficially the lineage is held to stretch back much further. According to the tradition, the current *tulku* can voluntarily choose whether he wishes to be reincarnated. Once his physical body

4 All information here is taken from the fourteenth Dalai Lama's official pronouncement on the nature of reincarnation, first published in 2011, and available online at http://www.dalailama.com/messages/statement-of-his-holiness-the-fourteenth-dalai-lama-tenzin-gyatso-on-the-issue-of-his-reincarnation.

dies, the search for the new *tulku* (his reincarnation) begins. Various procedures for identifying and recognizing a new *tulku* are approved. To quote from the current Dalai Lama's pronouncement on the topic:

> *After the system of recognizing Tulkus came into being, various procedures for going about it began to develop and grow. Among these some of the most important involve the predecessor's predictive letter and other instructions and indications that might occur; the reincarnation's reliably recounting his previous life and speaking about it; identifying possessions belonging to the predecessor and recognizing people who had been close to him. Apart from these, additional methods include asking reliable spiritual masters for their divination as well as seeking the predictions of mundane oracles, who appear through mediums in trance, and observing the visions that manifest in sacred lakes of protectors like Lhamoi Latso, a sacred lake south of Lhasa.*
>
> *When there happens to be more than one prospective candidate for recognition as a Tulku, and it becomes difficult to decide, there is a practice of making the final decision by divination employing the dough-ball method (zen tak) before a sacred image while calling upon the power of truth.*

When a new *tulku* is recognized, they are taken in by disciples of their predecessor and trained to keep up certain traditions and teachings. In this way their identity as the *tulku* is reinforced and sustained.

To those who do not share the system of religious beliefs, this can look like a bizarre practice. And to be absolutely clear, we do not discuss it in order to convince the reader of the Buddhist conception of mind, body, and reincarnation. We discuss it because it provides a case study of a human society in which there is widespread belief in the possibility of personality transfer, and because it illustrates how important processes of social construction are to the transference of personal identity. The current Dalai Lama *believes himself* to share an identity with an historical lineage due to social events and practices that have *caused* him to have certain beliefs and memories, to accept certain truths about his personal narrative, and to think about his relationship to the world in a particular way. From a first-person perspective, his identity (the continuing subject that he experiences himself to be) is causally socially constructed. If he did not grow up in Tibetan society, with the set of religious beliefs that it has, and if he was not taken in by the disciples of the thirteenth Dalai Lama at an early age and taught the beliefs and traditions of the *tulku*, he would not have the first-personal identity that he purports to have. Furthermore, from a third-person perspective, his social role, spiritual rights and responsibilities, is *constituted* by a network of collective beliefs.

The society around him believes that he shares a certain identity with the Bodhisattva of Compassion and so that is the identity that he is constructed as having within Tibetan society. This identity does not float free of all biological or psychological conditions of identity. The factors that are used to identify the next *tulku* make this much clear since they appeal to some such conditions. But the social factors are what ultimately prove decisive. They causally create and constitutively sustain the identity in that society.

PTSH will rely upon a very different metaphysical understanding and technical process of personality transfer, of course. But our argument is that social construction will be just as important to all instances of PTSH as it is to Buddhist reincarnation. We can see this by further spelling out the analogy between the two cases. Before any PTSH procedure there would be a biological human, after the procedure there would be a synthetic human, and there would be some attempt to claim that the biological human and the synthetic human shared the same personal identity. By analogy, then, the dead monk in the case of the Dalai Lama is like the biological human in a case of PTSH and the child is like the synthetic human. The reasons to claim that the previous Dalai Lama and the chosen child share a common identity look spurious at first, at least from the point of view of outsiders to this culture such as the authors. However, retrospectively, a compliant child will learn to behave in a manner consistent with expectations based on memories of the deceased monk's personality, providing *a posteriori* justification. Interestingly, just as in the case of a synthetic human, he would receive memories from the biological human though not through having actually lived those experiences in the synthetic body. The child will have some episodic memories from the life of the deceased monk that will have been passed to the child through word of mouth and thereafter internalized.

Why does personal identity function like this? We now describe the mechanisms that we believe are at work. The social construction of identity from the third-person perspective is easiest to explain. It is quite clear that certain properties of identity are constitutively constructed. The example of the Prime Minister's social identity is an obvious example of this. But what happens in this case happens to all of us, all the time. Certain features of our identity — our recognized relationships with others, our jobs, our wealth, our qualifications, and so on — are dependent upon collective belief. That collective belief is influenced by a variety of factors. For instance, recognizing that someone has a particular qualification is likely to be influenced by whether

or not they have a set of skills, but ultimately it is the collective belief that matters.

The social construction of identity from a first-person perspective is harder to explain. We argue that that it results from a combination of how our brains work and how cognition gets distributed between our brains and the environments in which we live. In other words, we argue that the socially constructed nature of our first-personal identities is a function of the embodied and distributed nature of cognition (Kirsh, 2006; Clark, 2008; Menary, 2007). A human's perceptual and cognitive abilities develop through life, and each human continually updates a highly compressed representation of their subjective history, i.e. their memories. The approximate nature of this memory compression is that raw sensations are grouped together to form perceptions, which are in turn grouped to form concepts, and so on hierarchically; concepts are the units which are linked together in memory (and which can be unpacked into perceptions and sensations if necessary to probe their meaning). The perceptions that humans have, that relate to their own body, form a conceptual grouping which we label 'self'. The human also interacts and communicates extensively with other humans and thereby forms concepts of those other humans. A 'person' is the conceptualization of a particular human, whether 'self' or 'other'. By extension, a personal identity consists in the set of concepts that relate to a particular person; importantly, this set of concepts resides not in one place but in the memories of both the human that the person is a conceptualization of, and of all their peers. This set of humans each have their own memories which relate to the person, and this entire body of information forms part of the personal identity.

Furthermore, some of the cognitive capacities and memories which constitute a personal identity are not purely held in the brain of the human to which it relates. We rely on networks of (socially constructed) cognitive artefacts to sustain our abilities to think and remember (Norman, 1991; Heersmink, 2013). These networks of cognitive artefacts include other people. For instance, a person may forget something that they had done or experienced, but then be reminded of it by one of their peers. The memory of the act was at one point present in the brains of both humans, then disappeared from the brain of the actor, but persisted in the brain of the peer, to be copied again to the brain of the actor at a later point (typically through verbal

transmission). Likewise memories can be temporarily offloaded into external media, e.g. photos. More generally, bodily form[5] and the environment can act as substrates for information relevant to personal identity. Thus we can think of a personal identity as a continually evolving set of information, which is stored disjointly across a set of different substrates, typically with the human body in question as the hub. In an extreme case of retrograde amnesia, all of the memories stored in the brain of the victim become inaccessible, yet the personal identity often remains linked to that human through the support of peers, who, with extensive effort, can help the victim to reassume their identity (Scott Bolzan is one such case in point — Bolzan, Bolzan and Rother, 2011). If, however, peers are not available, then the link between the human and their prior personal identity is broken and a new personal identity must be forged (Benjaman Kyle is a case in point — Wikstrom, 2011).[6] In addition to this, false memories can be as relevant to personal identity as true ones. If a person is falsely convicted of a crime, it affects how they are perceived and treated by their peers and this inevitably affects their own subsequent behaviour and self-image. A related point is that personal identities can be disjunctive, with contradictory memories about the person held simultaneously by different peers.

On top of this, personal identities can evolve in the absence of the human on which they are based. When a human dies, the personal identity suffers loss of the self-concept and related memories held by the human themself, but otherwise persists in all the memories that were external to the human. This personal identity can continue to evolve for a while, even in the absence of the cognitive processes of the human because, when peers communicate to each other about the person, the total body of information is modified. This process is obvious, for example, in the vilification of Jimmy Savile by the British public since his death (BBC, 2013). At the other extreme, when a pregnancy occurs, the parents-to-be go through preparations prior to the arrival of the baby. These may involve naming the child,

5 An example of bodily form could be a scar, which, when seen by the individual, brings to their mind the incident that caused the scar.

6 Wiley (2014, chapter 5) contains an extended discussion of the metaphysics of mind in the context of PTSH. His theory gives primary importance to the brain as the substrate that constitutes the mind, but his theory allows for the possibility of two physical humans constituted one mind, and highlights how brain states and external states combine to constitute mind states.

creating clothes and space for them, and informing an extended circle of peers. In so doing, a personal identity is created even before the baby has any real world experience or self-knowledge. If a miscarriage occurs then the parents can suffer grief; in this case there is a real sense in which a person has died, although there is only a tenuous connection between the person (a mental construct in the minds of the parents) and the physical fetus.

This social construction of identity can have odd repercussions. There is a standard assumption that a single personal identity relates to a single human body. However, this assumption breaks down in some unusual cases. We have already discussed the case of the Dalai Lama. We briefly describe two others here, starting with conjoined twins. Abigail and Brittany Hensel are perhaps a paradigm case (Pihlaja, 2008); they have two heads but one torso and one pair of arms and legs, and so their overall form is similar to that of a single human. They share some but not all internal organs and their nervous systems join from the lower spine downwards. They show some signs of acting as a single person; for example they can cooperate to perform complex real-time control tasks like playing basketball and driving a car with apparent ease, often with no need for verbal communication. They also sometimes write emails in the first person, as if they were a single person, when they both agree on what to say. In most other respects though they have different personality traits. Since there would be some grounds to consider them as a single person (single body; coordinated behaviour; sometimes speaking as one), a decision has been made (perhaps implicitly) by them and their peers to the contrary.

Another unusual case is that of dissociative identity disorder (DID). As an example, Kim Noble (Weitz, 2006; Mitchison, 2011) had DID from an early age, but it went undiagnosed or misdiagnosed for an extended period; in this time, peers observed that she had delusions, memory problems, erratic behaviour, etc. Once she was diagnosed with DID, this was a perspective from which her actions made more sense; there are several identities, seemingly with very clear separation of memories, time-sharing a single body, with abrupt changes from one identity to another. Interestingly, it was the act of psychologists in diagnosing her that allowed a redrawing of the boundaries leading to recognition of the multiple personal identities for which the body of Kim Noble is the substrate. None of Kim's personal identities

made this leap of understanding on their own, and some continue to deny it.[7]

Normally the association of a personal identity to a particular body is an automatic mental process, but both of the above cases help to illustrate that these associations result from decisions made with the help of social peers.

Based on these considerations, we conclude that significant elements of personal identity are socially constructed, both in the causal and constitutive senses of construction. This is not to say that identity does not depend on psychological or other conditions, but the importance of these conditions is often filtered through the social ones. Consequently, whether the synthetic human that results from a transferral process ultimately believes themselves to share an identity with a biological human, and is taken by society to share that identity, will depend on whether they are socially constructed as having that identity. This is what ultimately matters because, as long as people are willing to undergo PTSH in the hopes of survival — and we have every reason to expect that people will take bets on scientifically fanciful processes in the hopes of overcoming death[8] — there will be synthetic humans. The identities that those synthetic humans have will then depend on social factors.

6. Improving the Credibility of PTSH

The view of personal identity as a social construct has practical implications for approaches to PTSH. We close by reviewing these implications. First, there are no absolute criteria for what would constitute a successful transfer; rather, success would be judged by the maintenance of distributed cognitive frameworks and the willingness of peers to believe that transfer had occurred. Many factors could influence this willingness. Those who would like some form of PTSH to become a reality therefore have two paths open to them: they could work to solve some of the myriad technical problems that would need to be solved, for example with constructing suitable synthetic substrates; alternatively they could engage in promotion of the idea of PTSH with a view to increasing public acceptance. Concentrating on the former, there is a question of which approach to work towards.

---

7 Some argue that DID is iatrogenic rather than naturally occurring. We take no view on the matter here.

8 People choosing to enter cryonic storage upon death is a case in point.

Bamford (2012, Section 7) gave a preliminary discussion of some pros and cons of the various approaches. One might choose based on estimated technical feasibility, but it's fair to say that all of the proposed procedures summarized above are currently far beyond our expertise and there are many unsolved problems which may ultimately prevent any or all of these approaches from coming to fruition. The view we have presented suggests an alternative way to decide which procedures to work towards based on which are more likely to engender belief that transfer would occur; in other words, which factors would make PTSH more socially acceptable. In this section we discuss the factors we consider important.

6.1. Quantity and Quality of Information Transferred

Information relevant to the person which is held in the biological human must be moved to a synthetic human. It's trivially true that the more information transferred, the more credible the synthetic human would be as the new substrate for the personal identity, since loss of information would create results such as memory loss, cognitive differences, or, depending on details of the approach, general functional failure. Capturing this information is the technical aspect which has received the most attention from proponents of PTSH up to now. Hayworth (2012, p. 7), for example, posits that '…a particular human being's unique "software" consists of a discrete set of production rules, declarative memory chunks, and perceptual and motor memories…', and looks to the detailed structure of the brain's neural network as the carrier of this information. Rothblatt (2007; 2012) and Bainbridge (2009; 2012) instead focus on aspects of the behaviour of a person, such as the way they answer psychological quizzes, as a carrier of information with which a new substrate could be parametrized. These approaches would yield greatly differing types of information. The former would enable bottom-up reconstruction, starting from simple elements of a nervous system and building towards a functioning synthetic human. The latter would facilitate top-down reconstruction, starting from behaviours of a biological human and imposing these on a generic substrate. It remains to be seen which might be more useful in engendering credibility.

6.2. Similarity of Form and Function between Substrates

Some authors, e.g. Moravec (1988), point out that the robots of the future would not be constrained to resemble humans or any biological

life form. However, it would be easier to believe that a personal identity had been transferred if the synthetic substrate had the same form, physical capacities, and even distinguishing features as the biological human it were replacing. The work of Ishiguro in making robotic duplicates of particular people (Miyake *et al.*, 2011) is relevant in this respect. Likewise for function, if a synthetic substrate were capable e.g. of performing perception through standard human modalities, performing human-like cognition, understanding social contexts, and producing emotionally appropriate behaviour, such abilities would help to give a baseline of human-like behaviour over which the unique characteristics related to a personal identity could be imposed. Clearly, if a synthetic substrate were virtual in nature, as in the visions of some proponents of PST, this would be a fundamental change in form which would affect every aspect of the person's life, including the ways in which they could interact with peers.

6.3. Temporal Continuity of the Person

Some approaches to PTSH involve cryonic suspension, plastination, or some other long-term storage of information relevant to a deceased person, with the intention that the person could be reconstructed at some unspecified point in the future. Following our assertions above, their personal identity continues to evolve in the period in which they cease to act in society; they become known as deceased (heretofore understood to be an irreversible change); and their relationships with friends, family, coworkers, clients, etc. break down, for obvious reasons. Society continues to evolve separately from them, leading to a gradual loss of the context in which their life was rooted. If information stored from the biological human were then reanimated in a new synthetic substrate, the reintegration with those parts of the personal identity held in the memory of peers would be more difficult the more time had passed because the information would be more disjunctive. If a procedure provided either immediate or gradual transition of information between substrates, then the loss of credibility and other problems of integration that might be caused by a defunct period could be avoided. Hayworth (2010a) labelled as bad philosophy the idea that brain death would lead to the death of the person even if it could subsequently be reversed. We don't disagree with this labelling as bad philosophy but we would argue that if the idea were widely held in society it would be self-fulfilling and thus could not be so easily dismissed.

6.4. Spatial Continuity of the Body

Although the body-identity view of personal identity is philosophically problematic, it accords to a generally held intuitive view that a personal identity relates to a specific body. If the transition from biological to synthetic human were achieved by the deactivation, destruction, or otherwise the death of one, and the separate construction of the other, then we could say that the body which acted as the substrate underwent a spatial discontinuity. If instead the transition were achieved by a gradual process of replacement of body parts (including nervous system parts) with synthetic substitutes, then incredulity coming from intuitions about bodily continuity could be lessened.

6.5. Abruptness of Changes

This partly overlaps with considerations of spatial continuity and of similarity of form and function. Humans change physically through time and their related personal identities also evolve. If you fell out of touch with a friend and then met them again after 10 years, it should be no surprise that they looked different, had changed interests and attitudes, and had lost some memories of your previous interactions. If, however, the same changes occurred to them from one day to the next, it would seem pathological, and may even draw their identity into doubt. Part of the promise of passing to a synthetic substrate would be the possibilities of change that might be offered by the new substrate (as in e.g. the visions of Moravec, 1988). Allowing some change to occur as part of a PTSH procedure may also ease the technical constraints of the procedure, since it may allow a reduction in the quality or quantity of information that must be passed between substrates. However, where changes are necessary or desirable, allowing them to occur gradually while the person's social relationships continue to evolve should increase the credibility of continuity of personal identity. Goertzel (2012) has tried to formalize the idea of smoothness of change as it relates to continuity of personal identity; they have placed emphasis on the self-intelligibility of the person, whereas we would include criteria based on the observations of peers.

7. Conclusion

We have argued that 'personal transfer to a synthetic human (PTSH)' is an appropriate collective label for the set of hypothetical procedures

which currently go under the labels ‘mind uploading’, ‘substrate independent minds’, and ‘whole brain emulation’. This is based on a recognition that personal identity is the target for transfer which is desired by proponents of these ideas, and that where ‘mind’ is considered a target for transfer it is conflated with personal identity. We have argued for a definition of personal identity as a social construct. Specifically, we define a ‘person’ as a concept that a human forms regarding either itself (first-person perspective) or another human (third-person perspective), which is linked to memories of that particular human, and by extension as the set of all such concepts held about a particular human by the human themself and by all their peers, together with the memories which relate to that human. We have explored some consequences of this definition: personal identity consists partly of information which is not stored in the brain of the human but in the brains of other humans and in the wider environment (including the body); false memories can be as relevant to personal identity as true ones and personal identities can consist of contradictory information; personal identity can evolve in the absence of the human; although human bodies and personal identities normally have a one-to-one correspondence, one-to-many and many-to-one correspondences also exist. We have argued that the tradition of reincarnation practised by Tibetan Buddhists is an example of personal identity transfer between biological humans, which has been made to work by belief without relying on any technological interventions. This example suggests that PTSH could be possible but its success would be ultimately determined by social factors. Finally, we have discussed which features of proposed procedures for PTSH are likely to influence their credibility. These are: quantity and quality of information transferred; similarity of form and function between substrates; temporal continuity of the person; spatial continuity of the body; and smoothness of changes.

References

Bainbridge, W. (2009) Religion for a galactic civilization 2.0, *Institute for Ethics and Emerging Technologies*, [Online], http://ieet.org/index.php/IEET/more/bainbridge20090820/ [15 Jun 2017].

Bainbridge, W. (2012) Whole personality emulation, *International Journal of Machine Consciousness*, **4** (1), pp. 159–175.

Bamford, S. (2012) A framework for approaches to transfer of a mind’s substrate, *International Journal of Machine Consciousness*, **4** (1), pp. 23–34.

BBC (2013) *Jimmy Savile Scandal*, [Online], http://www.bbc.co.uk/news/uk-0026910 [12 Jan 2013].

Bolzan, S., Bolzan, J. & Rother, C. (2011) *My Life, Deleted: A Memoir*, San Francisco, CA: HarperOne.

Chalmers, D. (2014) Uploading: A philosophical analysis, in Blackford, R. & Broderick, D. (eds.) *Intelligence Unbound: The Future of Uploaded and Machine Minds*, Hoboken, NJ: Wiley-Blackwell.

Clark, A. (2008) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford: Oxford University Press.

Deca, D. (2012) Available tools for whole brain emulation, *International Journal of Machine Consciousness*, **4** (1), pp. 67–86.

Diaz-Leon, E. (2013) What is social construction?, *European Journal of Philosophy*, **23** (4), pp. 1137–1152.

Goertzel, B. (2006) *The Hidden Pattern: A Patternist Philosophy of Mind*, Irvine, CA: BrownWalker Press.

Goertzel, B. (2012) When should two minds be considered version of one another?, *International Journal of Machine Consciousness*, **4** (1), pp. 177–185.

Goertzel, B. & Ikle, M. (2012) Introduction to special issue, *International Journal of Machine Consciousness*, **4** (1), pp. 1–3.

Hacking, I. (1999) *The Social Construction of What*, Cambridge, MA: Harvard University Press.

Haslanger, S. (2003) Social construction: The 'debunking' project, in Schmitt, F. (ed.) *Socializing Metaphysics: The Nature of Social Reality*, Lanham, MD: Rowman & Littlefield.

Hauskeller, M. (2012) My brain, my mind, and I: Some philosophical assumptions of mind uploading, *International Journal of Machine Consciousness*, **4** (1), pp. 187–200.

Hayles, N.K. (1999) *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, Chicago, IL: University of Chicago Press.

Hayworth, K.J. (2010a) Killed by bad philosophy: Why brain preservation followed by mind uploading is a cure for death, *Brain Preservation*, [Online], http://www.brainpreservation.org/content-2/killed-bad-philosophy/ [15 Jun 2017].

Hayworth, K.J. (2010b) Open letter to the medical, scientific, and government communities regarding brain preservation, *Brain Preservation*, [Online], http://www.brainpreservation.org/wp-content/uploads/2015/08/Open-Letter-On-Brain-Preservation.pdf [15 Jun 2017].

Hayworth, K.J. (2011) Electron imaging technology for whole brain neural circuit mapping, *International Journal of Machine Consciousness*, **4** (1), pp. 87–108.

Heersmink, R. (2013) A taxonomy of cognitive artifacts: Function, information, and categories, *Review of Philosophical Psychology*, **4** (3), pp. 465–481.

Hongladarom, S. (2011) Personal identity and the self in the online and offline world, *Minds and Machines*, **21**, pp. 533–548.

Hopkins, P. (2012) Why uploading will not work, or the ghosts haunting transhumanism, *International Journal of Machine Consciousness*, **4** (1), pp. 229–243.

Hughes, J. (2011) Contradictions of the enlightenment: Liberal individualism versus the erosion of personal identity, *Institute for Ethics and Emerging Technologies*, [Online], https://ieet.org/index.php/IEET2/more/hughes20111119 [15 Jun 2017].

Kirsh, D (2006) Distributed cognition: A methodological note, *Pragmatics & Cognition*, **14** (2), pp. 249–262.

Koene, R. (2011) Pattern survival vs. gene survival, *Kurzweilai.net*, [Online], http://www.kurzweilai.net/pattern-survival-versus-gene-survival.

Koene, R. (2012a) Fundamentals of whole brain emulation: State, transition and update representations, *International Journal of Machine Consciousness*, **4** (1), pp. 5–21.

Koene, R. (2012b) Experimental research in whole brain emulation: The need for innovative in-vivo measurement techniques, *International Journal of Machine Consciousness*, **4** (1), pp. 35–65.

Kompridis, N. (2009) Technology's challenge to democracy: What of the human?, *Parrhesia*, **8**, pp. 20–33.

Kurzweil, R. (2000) *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, New York: Penguin.

Locke, J. (1690/1975) *An Essay Concerning Human Understanding*, Oxford: Oxford University Press.

Mallon, R. (2008) Naturalistic approaches to social construction, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*, [Online], https://plato.stanford.edu/entries/social-construction-naturalistic/.

Martins, N.R.B., Erlhagen, W. & Freitas, R.A. (2012) Non-destructive whole-brain monitoring using nanorobots: Neural electrical data rate requirements, *International Journal of Machine Consciousness*, **4** (1), pp. 109–140.

Menary, R. (2007) *Cognitive Integration: Mind and Cognition Unbounded*, London: Palgrave Macmillan.

Mitchison, A. (2011) Kim Noble: The woman with 100 personalities, *Guardian*, 30 Sept 2011, [Online], http://www.guardian.co.uk/lifeandstyle/2011/sep/30/kim-noble-woman-with-100-personalities.

Miyake, N., Ishiguro, H., Dautenhahn, K. & Nomura, T. (2011) Robots with children: Practices for human–robot symbiosis, *HRI '11 Proceedings of the 6th International Conference on Human–Robot Interaction*, DETAILS.

Moravec, H. (1988) *Mind Children*, Cambridge, MA: Harvard University Press.

Norman, D. (1991) Cognitive artifacts, in Carroll, J.M. (ed.) *Designing Interaction: Psychology at the Human–Computer Interface*, Cambridge: Cambridge University Press.

Olson, E.T. (2007) *What Are We? A Study in Personal Ontology*, New York: Oxford University Press.

Oto, B. (2012) Seeking normative guidelines for novel future forms of consciousness, *International Journal of Machine Consciousness*, **4** (1), pp. 201–214.

Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press.

Phillips, W. (2000) *Extraordinary Future*, [Online], http://www.mind.ilstu.edu/curriculum/listByAuthor.php.

Pihlaja, R. (2008) *Joined for Life: Abby and Brittany Turn 16*, (Film), Advanced Medical Productions Inc. USA.

Rothblatt, M. (2007) *Cyberev Project*, [Online], http://www.cyberev.org.

Rothblatt, M. (2012) The Terasem mind uploading experiment, *International Journal of Machine Consciousness*, **4** (1), pp. 141–158.

Sandberg, A. & Bostrom, N. (2008) *Whole Brain Emulation: A Roadmap*, Technical Report, Future of Humanity Institute, Oxford University, [Online], www.fhi.ox.ac.uk/reports/2008-3.pdf.

Schneider, S. (2009) Future minds: Transhumanism, cognitive enhancement and the nature of persons, in Ravitsky, V, Fiester, A. & Caplan, A.L. (eds.) *The Penn Center Guide to Bioethics*, New York: Springer Publishing Co.

Weitz, K. (2006) Kim Noble: A woman divided, *Independent*, 27 Aug 2006, [Online], http://www.independent.co.uk/news/people/profiles/kim-noble-a-woman-divided-413223.html.
Wikstrom, J. (2011) *Finding Benjaman*, (Film), PUBLISHER.
Wiley, K. (2014) *A Taxonomy and Metaphysics of Mind-Uploading*, Seattle, WA: Humanity+ Press and Alautun Press.

Paper received March 2017; revised June 2017.