The Informational Nature of Personal Identity

Luciano Floridi[1, 2]

[1]Research Chair in Philosophy of Information and GPI, University of Hertfordshire; [2]Faculty of Philosophy and IEG, University of Oxford.

Address for correspondence: Department of Philosophy, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK; l.floridi@herts.ac.uk

Abstract

In this paper, I present an informational approach to the nature of personal identity. In section one, I use Plato's famous metaphor of the chariot to introduce a specific problem regarding the nature of the self as an informational multiagent system: what keeps the self together as a whole and coherent unity? In section two and three, I outline two branches of the theory of the self: one concerning the *individualisation* of the self as an entity, the other concerning the *identification* of such entity. I argue that both presuppose an informational approach, defend the view that the *individualisation* of the self is logically prior to its *identification*, and suggest that such *individualisation* can be provided in informational terms. Hence, in section four, I offer an informational *individualisation* of the self, based on a tripartite model, which can help to solve the problem of the chariot. Once this model of the self is outlined, in section five I use it to show how ICTs may be interpreted as technologies of the self. In section six, I introduce the concept of "realization" (Aristotle's *anagnorisis*) and support the rather Spinozian view according to which, from the perspective of informational structural realism, selves are the final stage in the development of informational structures. The final section seven briefly concludes the article with a reference to the purposeful shaping of the self, in a shift from egology to ecology.

Introduction

Questions about personal identity and the nature of the self are as old as philosophy,[1] so one may suspect that nothing new could sensibly be said about the topic. Such suspicion, though reasonable, would be justified only in part. In the last decade or so (Turkle 1995), a new area of investigation into the nature of personal identity has begun to emerge, due to the dramatic evolution of information and communication technologies (ICTs) and their widespread impact on our lives (Floridi 1995, 2010). Today, we increasingly acknowledge the importance of a common yet unprecedented phenomenon, which may be described as the construction of personal identities in the infosphere. Human life is quickly becoming a matter of *onlife* experience, which reshapes constraints and offers new affordances in the development of our identities. Elsewhere,[2] I have suggested that such a phenomenon is part of a wider trend, a fourth revolution in our self-understanding, brought about by computer science and ICT applications, after the Copernican, the Darwinian and the Freudian ones. In this article, I intend to explore the foundations of the construction of personal identities, by developing an informational analysis of the self.[3] The broader thesis I shall defend is that ICTs are, among other things, egopoietic technologies or technologies of self construction, significantly affecting who we are, who we think we are, who we might become, and who we think we might become. Such thesis is articulated and supported through the following steps.

In section one, I shall rely on Plato's famous metaphor of the chariot in order to introduce a specific problem regarding the nature of the self as an informational multiagent system: what keeps the self together as a whole and coherent unity? This question may be addressed from two perspectives. One is *synchronic* and focuses on what may constitute the self as a particular whole unity, continuously existing and coherently behaving at any given time. The other is *diachronic* and focuses on what may enable the self to remain that unity, or simply itself, at different times and through changes. Following this distinction, in section two and three I shall quickly outline two branches of the theory of the self, to be labelled *egology* for short: one

[1] In writing this article, I relied especially on (Martin and Barresi 2006), (Perry 2008), and (Sorabji 2006).
[2] On the information turn as a fourth revolution in our self-understanding, after the Copernican, the Darwinian and the Freudian, see (Floridi 2007, 2008a, 2011a).
[3] The informational analysis connects this article to previous work on the ontological interpretation of informational privacy, see (Floridi 2005b, 2006).

concerning the *individualisation* of the self as an entity (no substantialism, essentialism, or dualism is presupposed), the other concerning the *identification* of such an entity. I shall argue that both presuppose an informational approach, defend the view that the *individualisation* of the self is logically prior to its *identification*, and suggest that such *individualisation* can be provided in informational terms. In section four, I shall then offer an informational *individualisation* of the self based on a tripartite model, illustrated in terms of a three-membrane description: the corporeal, the cognitive, and the conscious. This 3C model of the self helps to tackle the problem of the chariot. Once it is outlined, in section five I shall use it to show how ICTs may be interpreted as technologies of the self, by illustrating how they affect each membrane. The informational interpretation of the self would be incomplete without the inclusion of a reflection on the very understanding of the self by the self. Such "self-understanding" is provided in section six, where I shall borrow Aristotle's concept of *anagnorisis* ("realization") in order to support the rather Spinozian view according to which, from the perspective of informational structural realism (Floridi 2008b, 2011b), selves are the final stage in the development of informational structures, for they are the semantically structuring structures conscious of themselves. The final section seven briefly concludes the article with a reference to the purposeful shaping of the self, in a shift from egology to ecology.

1. Plato and the Problem of the Chariot

In one of the most famous passages in the history of philosophy, Plato compares the soul – what in this article will be referred to as the self – to a chariot:

> We will liken the soul to the composite nature of a pair of winged horses and a charioteer. [...] the charioteer of the human soul drives a pair, one of the horses is noble and of noble breed, but the other quite the opposite in breed and character. Therefore in our case the driving is necessarily difficult and troublesome. *Phaedrus* 246a - 254e

The tripartite analogy is too well known to deserve any explanation, but two aspects of it may be highlighted here, for they nicely introduce both the "engineering" approach adopted in the following pages, and the key problem on which I shall focus.

First, the approach. Plato quite literally interprets the self as a multiagent system (MAS), and not just any MAS, but one that has a significantly technological nature. Look carefully and you will see that the three agents are not three sides of a

triangle, “three men on a boat”, a master and two slaves, or a family of two parents and a child. They are three components in a complex, engineered artefact, and one that was fairly advanced for the time. Thus, Plato’s technological analogy of the multiagent chariot is interesting both because it facilitates the application of a wealth of interesting results to the analysis of the self, already available in the literature on MAS (Wooldridge 2009), and because it invites a shift from a phenomenological or descriptive approach to the self to a constructionist or design-oriented approach, one that considers what it means to create (or at least what it means for something to constitute) such a chariot or multiagent system. It is easy to realise, for example, that some of the classic challenges in the engineering of MAS (Bond and Gasser 1988; Sycara 1998) – such as communication, coherence, rationality, successful interaction with the environment, coordination and collaboration with other agents, just to mention the most obvious – are just AI translations of classic issues in the philosophy of the self. Still from a design perspective, upbringing, training, education, social and political practices and norms may easily be interpreted as selves-engineering techniques, as Plato already knew, and any virtue ethics rightly assumes. The comparison could be extended, and I shall briefly return to this self-engineering process in the conclusion. At the moment, let me highlight the second aspect, which, quite surprisingly, seems to have been overlooked by the vast literature on the Platonic analogy. A difficult question posed by any multiagent analysis of a system, be this an engineered artefact, a society of agents (Minsky 1986), or a biological self, is: what makes such a MAS a coherent unity and source of actions, and keeps it as such? The question may not immediately strike one as difficult in engineering contexts, where we build the MAS in which we are interested as units, but even there the problem soon becomes pressing once we start considering slightly more complex scenarios, in which agents temporarily coordinate their actions and collaborate to achieve specific goals (e.g. a rowing team). In biology, the study of multi-cellular organisms made up of specialized tissues and organs already shows the complexity of the problem. In philosophy, one appreciates its difficult nature as soon as one recognises in it an instance of the infamous problem of Theseus’ ship. If one of the two horses is replaced, is it the same soul? And what happens if the charioteer decides to dismount the chariot and leave the horses to their destiny? More seriously, it seems plausible to assume that the MAS in question is constituted by its interacting and coordinated components and may not survive either their complete replacement or

their irrecoverable disappearance, but what about their evolution? Such questions help to clarify the fundamental challenge posed by the unity of the self. I shall refer to it as *the problem of the chariot* because it is the chariot and the tack that, in Plato's analogy, represent the fourth, hidden component that guarantees the unity and coordination of the system, thus allowing the self to be, persist and act as a single, coherent, and continuous entity in different places, at different times, and through a variety of experiences. It is the problem of the chariot that poses a serious challenge to any information-based theory of the self, as we shall see in the next two sections.

2. Egology and its Two Branches

Plato's interest in the theory of the self, or *egology*, was ethico-political and epistemological, not yet ontological. Therefore, his dialogues explore the life of the multiagent system (the tripartite self, the socially structured city), but leave the problem of the chariot philosophically (if not mythologically) untouched. It is mainly from Descartes onwards that the unity, identity, and continuity of the I, or self, as an entity become the subjects of an ontological investigation in their own right. It takes the Christian emphasis on the concept of individual person and then the long-term fading of a Christian answer to what an individual person is, to place egology at the centre of philosophical attention first, and then turn it into a source of problems. Once modern egology becomes an ontology of the self, two branches soon emerge. *Diachronic* egology, understood as an ontology of personal *identity*, concentrates on the problems arising from the *identification* of a self through time or possible worlds, progressively moving towards metaphysics. *Synchronic* egology, understood as an ontology of *personal* identity, deals with the *individualisation* of a self in time or in a possible world, thus placing itself at the heart of the philosophy of mind. For reasons that will become clear presently, in the rest of the article I shall focus only on synchronic egology. So let me devote the rest of this section to sketching the sort of approach that might be developed when dealing with diachronic egology informationally.

As it is well known, the literature on diachronic egology offers two main alternatives. *Endurantism* argues that a self is a three-dimensional entity that wholly exists at each moment of its history, and the same self exists at each moment. *Perdurantism* argues that a self is a four-dimensional entity constituted by a series of

spatial and temporal parts, somewhat like the frames of a film. In both cases, an ontology of the self is developed by presupposing some form of direct realism, according to which the model (description, theory, representation, analysis etc.) of the system (the referent of the model, in this case the self, the I, or whatever is intended by personal identity as a feature of the world) can be developed through a non-mediated access to the system in itself. Such presupposition may be justified, but is certainly open to question for all those who, like myself, are convinced that any system, the self included, is always accessed and hence modelled at a given level of abstraction or LoA.[4] This suggests an alternative approach, according to which the analysis of self "identity" (*a* is this) and "sameness" (this is the same *a* as that *a*) relations should be developed in terms of the relevant kinds of information (observables) that, once fixed, provide the referential framework required to satisfy the specific epistemic goals in question. If this is unclear, consider the following example. Whether a hospital transformed now into a school is still the same building seems a very idle question to ask, if one does not specify in which context and for which purpose the question is formulated, and therefore what the required observables are that would constitute the right LoA at which the relevant answer may be correctly provided. If the question is asked in order to get there, for example, then the relevant observable is "location" and the answer is yes, they are the same building. If the question is asked in order to understand what happens inside, then "social function" is the relevant observable and therefore the answer is obviously no, they are very different. The illusion that there might be a single, correct, absolute answer, independently of context, purpose and LoA, leads to paradoxical nonsense. Nor does the retort that some LoAs should be privileged when personal identities are in question carry much weight. For the same analysis holds true when the entity investigated is the young Saul, who is watching the cloaks of those who laid them aside to stone Stephen (Acts 7:58), or the older Paul of Tarsus, after his conversion. Saul and Paul are and are not the same person; the butterfly is and is not the caterpillar; Rome is and is not the same city in which Caesar was killed and that you visited last year; you are and yet you are not the same person who went there. It depends on the LoA, and this depends on the purpose for which, and the context in

[4] The reader unacquainted with the method of levels of abstraction in computer science might do worse than just imagining a LoA as an epistemic interface. The interested reader might wish to check (Floridi 2008c).

which the question is asked. Locke was right in urging us to be careful about the sort of question that one might ask about the same man, same substance, same soul, same consciousness, same set of memories etc., and also about the LoA that one is naturally led to privilege (the consciousness one), especially from a first-person perspective. He was wrong – indeed incoherent, for someone who acknowledged, correctly, not to know what substance might be in itself – in committing himself ontologically, when an informational (epistemological, for Locke) standpoint would have been sufficient. Identity and sameness relations are satisfied according to the LoAs adopted, and these, in turn, depend on the goals being pursued. This is not relativism: given a particular goal, one LoA is better than another, and questions will receive better or worse answers. The ship will be Theseus's, no matter how many bits one replaces, if the question is about legal ownership (try a Theseus trick with the taxman); it is already a different ship, for which the collector will not pay the same price, if all one cares about are the original planks. Questions about diachronic identity and sameness are really teleological questions, asked in order to attribute responsibility, plan a journey, collect taxes, attribute ownership or authorship, trust someone, authorise someone else, and so forth. Insofar as they are dealt with metaphysically (modally or not, it does not matter), they do not deserve to be taken seriously. For in a LoA-free context they make no sense (although it might be intellectual fun to play idly with them), exactly like it makes no sense to ask whether a point is at the centre of the circumference without being told what the circumference is, or being told the price of an item but not the currency in which it is given. It is not just the degree of confidence in the re-identification through time or possible worlds of someone as the same someone that is a matter of epistemology; it is the very process of identification and re-identification that needs to be conceptualised in a fully epistemological way, i.e. informationally, through a careful analysis of the information that is being required and hence needs to be made available to provide a reasonable answer, because

> That which has made the difficulty about this relation [sameness], has been the little care and attention used in having precise *notions* [i.e. *information*, my emphasis] of the things to which it is attributed. (Locke 1979), Book II, Chapter XXVII, §§ 27-30.

Let us now turn to the individualisation of the self.

3. Egology as Synchronic Individualisation

Before being able to establish, informationally (i.e., at the right LoA), whether this *a* is even approximately the same as that *a*, it seems that one needs to have some information about what this *a* is. Plato was right: you cannot look for something, let alone know whether you found it, unless you know what you are looking for. So, individualisation logically precedes identification. Of the many approaches that seek to characterise the nature of the self, two stand out as popular and promising for the task ahead: the Lockean one, according to which the identity of the self is grounded in the unity of consciousness and the continuity of memories; and the Narrative approach (Schechtman 1996), according to which a self is a socio- or (inclusive or) auto-biographical artefact. We have already encountered Locke in the previous section. Regarding the Narrative approach, the following passage elegantly illustrates its essential gist:

> But then, even in the most insignificant details of our daily life, none of us can be said to constitute a material whole, which is identical for everyone, and need only be turned up like a page in an account-book or the record of a will; our social personality is created by the thoughts of other people. Even the simple act which we describe as "seeing some one we know" is, to some extent, an intellectual process. We pack the physical outline of the creature we see with all the ideas we have already formed about him, and in the complete picture of him which we compose in our minds those ideas have certainly the principal place. In the end they come to fill out so completely the curve of his checks, to follow so exactly the line of his nose, they blend so harmoniously in the sound of his voice that these seem to be no more than a transparent envelope, so that each time we see the face or hear the voice it is our own ideas of him which we recognise and to which we listen. (Proust 1992), *Overture*.

We "identify" (provide identities) to each other, and this is a crucial (although not the only) variable in the complex game of the construction of personal identities, especially when the opportunities to socialise are multiplied and modified by new information technologies, as we shall see.

Now, in both cases, individuation – the characterization or constitution of the self – is achieved through forms of information processing: consciousness and memory are dynamic states of information, but so is any kind of personal or social narrative. So both the Lockean and the Narrative approach presuppose the existence of individual agents endowed with the right sort of informational skills. Hume saw this quite clearly, but was also aware that his account of the "informational" self

completely failed to explain its unity. The passage is famous but it is worth quoting at length while keeping in mind the problem of the chariot:

> […] having thus loosen'd all our particular perceptions [bits or streams of information separate from each others], when I proceed to explain the principle of connexion, which binds them together, and makes us attribute to them a real simplicity and identity; I am sensible, that my account [the bundle and then the commonwealth] is very defective […]. If perceptions are distinct existences, they form a whole only by being connected together. But no connexions among distinct existences are ever discoverable by human understanding. We only *feel* a connexion or a determination of the thought, to pass from one object to another. It follows, therefore, that the thought alone finds personal identity, when reflecting on the train of past perceptions, that compose the mind […].Most philosophers seem inclin'd to think, that personal identity *arises* from consciousness; and consciousness is nothing but a reflected thought or perception [information processing]. The present philosophy, therefore, has so far a promising aspect. But all my hopes vanish, when I come to explain the principles, that unite our successive perceptions in our thought or consciousness. […] In short, there are two principles, which I cannot render consistent; nor is it my power to renounce either of them, viz. *that all our distinct perceptions are distinct existences* and *that the mind never perceives any real connexion among distinct existences* [the infrastructure that keeps them together as a unity]. Did our perceptions either inhere in something simple and individual [the tack], or did the mind perceive some real connexion among them, [if there were a chariot] there would be no difficulty in the case. For my part, I must plead the privilege of a sceptic, and confess that this difficulty is too hard for my understanding. I pretend not, however, to pronounce it absolutely insuperable. Other, perhaps, or myself, upon more mature reflection, may discover some hypothesis, that will reconcile those contradictions. (Hume 2007), Appendix, §§ 20-21, vol. 1, p. 400.

In short: if the self is made of information (perceptions or narratives, or any other informational items one may privilege), then a serious challenge is to explain how that information is kept together as a whole, coherent, sufficiently permanent unity. If there is no narrator – and there cannot be, because the narrative theory of the self describes the narrator as the narrative, and presupposing a narrator would only shift the problem one step back – what prevents the narrative from being a completely random, incoherent and disjointed selection of miscellaneous bits of stories? The answer seems to be twofold. First, there is a blocking manoeuvre, which prevents us from biting the bullet: selves, if they are narratives, are coherent and unitary narratives, at least when dealing with healthy, ordinary selves. We owe this to Kant, who made a step forward by arguing convincingly (or at least so plausibly as to shift the burden of proof on the shoulders of those who disagree) that the unified coherence

of the information about the external world, synthesised by the epistemic agent, could be guaranteed only by the unity of the very agent's self that is its (of the information) source. So Kant's transcendental argument, in favour of the unity of the self, is a partial, epistemological solution to the ontological problem of the chariot, or the unity of the informational self. Yet it is only "partial" because, like all transcendental arguments, it is non-constructive, to use a mathematical distinction. At best, it shows that a specific characterization of the self as a whole unity of consciousness is the required condition of possibility for the meaningful coherence of the stream of empirical information generated by the agent. How such unity and coordination come to be there in the first place and have those features is not the issue addressed. It is the part of the question left unanswered. Kant is essentially arguing that the chariot and the tack must be there and have the features that they have in order for the MAS to work informationally as successfully it does, but he provides no further insight into how such unity arises or is reached in the first place, and then maintained. So we are still left with the problem: granted that the unity of the narrative or informational self and (perhaps) its crucial role in the delivery of a coherent experience of the world must be conceded, what generates it and keeps it together?

Following Kant, I too left such a question unanswered in the past. In (Floridi 2005b, 2006), I defended an informational interpretation of the self, arguing that each self should be conceptualised as being constituted by its information, thus understanding a breach of one's informational privacy as a form of aggression against one's personal identity. The thesis has its roots in the classic analysis by (Warren and Brandeis 1890), in which they argue (p. 33) that

> the right to privacy, as part of the more general right to the immunity of the person, [is] the right to one's personality.

Yet, the problem is that, if the flow of information (or Humean perceptions, or narrative elements) is no more than an aggregate, it must fail to form a coherent unity, let alone a conscious self, unless it is consistently and non-transiently bound together as a whole, but then the binding, that is, the problem of the occurrence and maintenance of the chariot, is precisely an instance of our recurring difficulty.

Clearly, it is going to be hard to tackle a problem that Plato, Hume and Kant left unsolved. We do have the advantage of coming after them and hence being able to learn from, and build upon, their work. Still, such advantage might come at a high

price, in terms of what plausible solutions are still viable. In the following section, I shall follow Sherlock Holmes' advice: having eliminated the impossible, whatever remains, however improbable, must be the truth. But the reader should know that I am aware that "Other, perhaps, upon more mature reflection, may discover some hypothesis, that will reconcile those contradictions" and that escaped my understanding.

4. A Reconciling Hypothesis: the Three Membranes Model

Kant was able to show that the unity of the self must be presupposed as the source that "unite[s] our successive perceptions in our thought or consciousness", to quote Hume once again. In this section, I shall suggest that such informational unity of the self may be achieved, or at least described, through a three-phase development of the self. The model I am going to propose is, I take it, biologically and informationally plausible, but it is, admittedly, somewhat figurative. I hope the reader will not object. On the one hand,

> To tell what it really is [the form of the soul, or the characterisation of the self] would be a matter for utterly superhuman and long discourse, but it is within human power to describe it briefly in a figure; let us therefore speak in that way. Plato, *Phaedrus*, 246a

On the other hand, the goal is ultimately that of explaining in what sense ICTs are egopoietic technologies and I hope that the model at least achieves this much.

The "reconciling hypothesis", to use Hume's terminology, that I wish to articulate is strategically simple, if a bit complicated in its details. Here it is. In the same way that organisms are initially formed and kept together by auto-structuring (i.e., auto-assembling and, within the assembled entity , auto-organising)[5] physical (henceforth *corporeal*) membranes, which encapsulate and hence *detach* (bear with me, more on this below) parts of the environment into biochemical structures that are then able to evolve into more complex organisms, selves too are the result of further encapsulations, although of informational rather than biochemical structures. The basic mechanism of encapsulation, detachment and internal auto-organization, I suggest, is the same, or at least we should take seriously the possibility that it might be the same from a minimalist perspective (Ockham's razor). If this is the case, then

[5] In the paper I use "auto-" instead of "self-" in order to avoid potential confusions whenever necessary.

selves emerge as the last step in a process of detachment from reality that begins with a corporeal membrane encapsulating an organism, proceeds through a cognitive membrane encapsulating an intelligent animal, and concludes with a consciousness membrane encapsulating a mental self or simply a mind. Of course, one may add as many mid-steps as required, yet these three – the corporeal, the cognitive and the consciousness or simply 3C – seem to be the main stations at which the train of evolution has called. Each step builds on the previous one (supervenience) and, at each step, more not less distance is placed between the entity and its environment. Each membrane is a defence of the structural integrity of what it encapsulates, against the surrounding environment. Of course, in moving from the corporeal, to the cognitive to the consciousness membrane, there is an increasing process of virtualisation. Yet, there is nothing metaphorical in this, as anyone acquainted with the concept of the virtual machine in computer science can readily appreciate. Indeed, it has become almost fashionable to compare the mind to a virtual machine,[6] even if, without some further theorising, the comparison only hides and fails to solve the usual *homunculus* problem. Nor is there any problem about each membrane being auto-poetically structured through auto-assembly and auto-organization: at each stage, local relations act on local building blocks to generate a new divide, within the old environment, between a new inside and hence a new outside. This is the general hypothesis. Let me now add some details about the model.

The three phases concern the evolution of organisms, then of intelligent animals and finally of self-conscious minds. Each phase contributes to the construction of the ultimate personal identity of the human organism in question.

Phase one: the physical membrane $\Longrightarrow$ organisms

The first phase begins in an environment in which there are not yet biotic structures. There are, however, physical structures, that is, patterns of physical data understood

[6] See for example the symposium in (Hayes et al. 1992) or the debate between (Densmore and Dennett 1999) and (Churchland 1999). To the best of my knowledge, Aaron Sloman has been the first to call attention to the computational theory of virtual machines as a way to model the mind, see (Sloman and Chrisley 2003) for a more recent statement. I agree with (Pollock 2007) that, in general, the whole approach seems a refined version of the sort of classic functionalism originally developed by (Putnam 1960). As such, "virtual functionalism" does not seem to be much more instructive than the old fashioned kind. For example, Pollock writes "If I am a virtual machine, which virtual machine am I? The proposal is that I am a virtual machine that cognizes. But there is more than one such virtual machine implemented on my body." Clearly it is the concurrence of machine-like processes that is "solved" by the virtualization of the machine itself, a gain that does not seem to be a substantive progress with respect to any alternative analysis in terms of multi-functionality. Unfortunately, virtual machines generate virtual problems about virtual minds virtually conscious.

as asymmetries or lacks of uniformities, e.g. lights, noises, or magnetic fields. Such data might be flowing around, but there are no senders or receivers yet. This might be seen as a stage when there are environmental data and patterns that might be exploitable as information by the right sort of agents for their purposes, before there is any kind of communication. We move from a pre-biotic to a post-biotic environment once some structures in the environment become encapsulated through a *corporeal membrane*. The encapsulation of part of the environment through a corporeal membrane allows the separation of the interior of a cell from the external world. This is the ontological function of the membrane, as a hardwired divide between the inside, the individual biotic structure, and the outside, the environment. Its negentropic function is to enable the organism to interact with the environment to its own advantage and withstand for as long and as well as possible the second law of thermodynamics. The epistemological function of the membrane is that of being selectively-permeable, thus enabling the cell a variety of degrees of inputs and outputs with respect to the environment. At this stage, data are transduceable physical patterns, that is, physical signals now seen as broadcasted by other structures in the environment, which are captured by the permeable membrane of the organism. The body is a barrier that protects the stability of the living system (physical homeostasis). A good example is a sunflower.

Phase two: the cognitive membrane $\Longrightarrow$ intelligent animals

We move from pre-cognitive to post-cognitive systems once data become encodable resources exploitable by organisms through some language broadly conceived (sounds, visual patterns, gestures, smells, behaviours etc.). This requires a *cognitive membrane*, which allows the encapsulation of data for processing and communication. The streams of data, which were before quantities without direction (scalars), broadcasted by sources not targeting any particular receiver (e.g. the sun generating heat and light, or the earth generating a magnetic field), acquire a direction, from sender to receiver (vectors), and an interpretation (e.g., noises become sounds interpreted as alarms). From now on, Shannon's classic communication model applies. The body becomes an interface and the cognitive membrane is a semi-hardwired (because configurable) divide between the cognitive system and its environment, that is, a barrier that further detaches the organism from its surroundings, and allows it to exploit data processing and communication in its fight

against entropy. The stability (cognitive homeostasis) now concerns the internal data within the system and their codification: memory and language. A good example is a bird on the sunflower.

Phase three: the consciousness membrane $\Longrightarrow$ self-conscious minds

The third phase is represented by the evolution of the consciousness membrane. We move from pre-conscious (aware) to post-conscious (self-aware) systems once data become repurposable information, including conventional meanings (e.g. sounds become a national anthem). The consciousness membrane is softwired (programmable). The body becomes the outside environment for an inside experience, and the stability now concerns the self within the system (mental homeostasis). To put it in Cartesian terms, the mind is indivisible, not because it cannot be divided (detached from itself) but because the division (detachment) does not generate two minds, but mere schizophrenia. This is why there is no further, healthy detachment of the self from the self, but only increasing degree of self-reflection. Once the self or I emerges, it appropriates and unifies what happens to the corporeal and cognitive levels as his or her own experiences.[7] A good example is a gardener watching the bird on the sunflower.

The 3C model just sketched helps us to deal with the problem of the chariot and, in so doing, it finally enables us to clarify why, and in what sense, ICTs are technologies of the self. Each membrane, and hence each step in the detachment of the individual from the world, is made possible by a specific, auto-reinforcing, bonding force. The corporeal membrane relies on chemical bonds and orientations. The cognitive membrane relies on the bonds and orientations provided by what is known in information theory as mutual information, that is the (measure of) the interdependence of data (the textbook example is the mutual dependence between smoke and fire). And, finally, the consciousness membrane relies on the bonds and orientations provided by semantics (here narratives provide plenty of examples), which ultimately makes possible a stable and long-lasting detachment from reality. At each stage, corporeal, cognitive and consciousness elements fit together in structures (body, cognition, mind) that owe their unity and coordination to such bonding forces.

[7] In (Floridi 2005a) I have defined this as the I before Mine hypothesis, or IBM.

The more virtual the structure becomes, the more it is disengaged from the external environment in favour of an autonomously constructed world of meanings and interpretations. The I emerges as a break with nature, not as a super connection with it. Such an "unnatural" break requires a collaborative and cumulative effort by generations through time. No individual can successfully rely just on a private semantics (what Wittgenstein calls private language). This is why a single human being needs to be embedded, at a very early stage of development, within a community, in order to grow as a healthy self-conscious mind: mere corporeal and cognitive bonds, in one-to-one interactions with the external environment, fail to give rise to, and keep together, a full self, for which language, culture and social interactions are indispensable. The problem of the chariot therefore may be solved only by taking into account all the bonding forces – physical, cognitive, and semantic – that progressively generate the unity of the self. As Hume discovered, by itself each of them is insufficient.

The 3C model as a solution of the problem of the chariot acquires further plausibility once we apply it to explain the impact of ICTs on the construction of personal identity. This is the topic of the next section.

5. ICTs as Technologies of the Self

If the self is made possible by the healthy development of all the three membranes, then any technology capable of affecting any of them is *ipso facto* a technology of the self. Already Plato, for example, acknowledged that humanity had changed because of the invention of writing. Now, ICTs are the most powerful technologies to which selves have ever been exposed. They induce radical modifications (a re-ontologisation) of the contexts (constraints and affordances) and praxes of self-poiesis, by enhancing the corporeal membrane, empowering the cognitive membrane, and extending the consciousness membrane. Let us have a quick look. The following examples are not meant to provide an exhaustive analysis but a variety of brief illustrations about embodiment, space, time, memory and interactions, and finally perception.

5.1 Embodiment: from Dualism to Polarism

We have seen that each membrane contributes to the construction of the self: the body, its cognitive functions and activities and the consciousness that accompany them are inextricably mixed together to give rise to a self and its personal identity. Diachronically, each membrane must be there for the others to occur. Yet this truism hides the fundamental fact that, once a membrane is in place, the particular inside that it detaches from the relevant outside becomes conceivably independent of the previous stages of development. It is correct to stress that there is no butterfly without the caterpillar, but insisting that once the butterfly is born the caterpillar must still be there for the butterfly to live and flourish is a conceptual confusion. There is no development of the self without the corporeal and the cognitive faculties, but once the latter have given rise to a consciousness membrane, the life of the self may be entirely internal and independent of the specific body and faculties that made it possible. While in the air, you no longer need the springboard, even if it was the springboard that allowed you to jump so high, and your airborne time is limited by gravity. Wittgenstein is right in saying that no private language may subsist without a public language, but once a public language is available, the speaker may throw away the public language (privatise it, as it were), like the famous ladder. This does not mean that the self requires no physical platform. Some platform (some data structure) is required to sustain the constructed self. And it does not mean that just any platform would do either. But it does open the possibility of a wider choice of platforms. Our culture, so imbued with informational concepts, finds the very idea of eterobodiment of the self, or the self as a cross-platform (not a-platform) structure, perfectly conceivable, witness the debate about mind uploading and body swap in the philosophy of mind. It is not the science fictional nature of the thought experiments that is interesting – in many cases it tends to be distracting and fruitlessly scholastic – but the readiness with which many seem to be willing to engage with them, because this is indicative of the particular impact that ICTs have had on how we conceptualise selves.

5.2 Space: the Detachment between Location and Presence

Through the phenomenon of telepresence, ICTs magnify (make more salient and increases) the distinction between presence vs. location of the self. A living organism

(e.g. a spider) is cognitively present only where it is located as an embodied and embedded information-processing system. A living organism aware of its information processes (e.g. a dog dreaming) can be present within such processes (e.g. chasing dreamed rabbits) while being located elsewhere (e.g. in the house). But a self, that is, a living organism self-aware of its own information processes (e.g. you) and its own presence within them, can choose where to be. The self, and mental life in general, is located in the brain but not present in the brain. Thus the locus of the self is the brain but the self is not present in the brain.

5.3 Time: the Detachment between Outdating and Ageing

ICTs increase the endurance effect, for in digital environments exactly the same self may be identified and re-identified through time. The problem is that the virtual may or may not work properly, it may be old or new, but it does not grow old; it outdates, it does not age. Nothing that outdates can outdate more or less well. On the contrary, the self ages and does so more or less well. The effect, which we have only started to experience and are still learning to cope with, is a chronological disalignment between the self and its online habitat, between parts of the self that age and parts that simply outdate. Asynchronicity is acquiring a new meaning.

5.4 Memories and Interactions: Fixing the Self

We have seen that memory plays a crucial role in the construction of personal identity. Obviously, any technology, the primary goal of which is to manage records, is going to have an immense influence on how individuals develop and shape their own personal identities. It is not just a matter of mere quantity; the quality, availability and accessibility of personal records may deeply affect who we think we are and may become. Until recently, the optimistic view was that ICTs empowered individuals in their personal identity DIY. The future is more sombre. Recorded memories tend to freeze the nature of their subject. The more memories we accumulate and externalise, the more narrative constraints we provide for the construction and development of personal identities. Increasing our memories means decreasing the degree of freedom we might enjoy in defining ourselves. Forgetting is also a self-poietic art. A potential solution, for generations to come, is to be thriftier with anything that tends to fix the nature of the self, and more skilful in handling new or refined self-poietic skills. Capturing, editing, saving, conserving, managing one's

own memories for personal and public consumption will become increasingly important not just in terms of protection of informational privacy, but also in terms of construction of one's personal identity. The same holds true for interactions, in a world in which the divide between online and offline is being erased. The onlife experience does not respect dimensional boundaries, with the result that, for example, the scope for naïve lying about oneself on Facebook is increasingly reduced (these days everybody knows if you are, or behave like, a dog online). In this case, the solution may lie in the creation of more affordances and spaces for self-expression and self-poiesis (e.g. Diaspora, the open source Facebook).

5.5 Perception: the Digital Gaze

The gaze is a composite phenomenon, with a long and valuable tradition of analyses (Lacan, Foucault, Sartre, Feminist theory). The idea is rather straightforward: the self observes "the observation of itself" by other selves (including, or sometimes primarily itself) through some medium. It should not be confused with seeing oneself in a mirror (ego surfing or vanity googling). It is rather comparable to seeing oneself as seen by others, by using a mirror ("what people see when they see me?"). In child development, the gazing phase is theorised as a perfectly healthy and normal stage, during which the individual learns to see her or himself by impersonating, for example, a chair ("how does the chair sees me?"), or simply placing her or himself in someone's shoes, as the phrase goes. The digital gaze is the transfer of such phenomenon in online environments. The self tries to see how others see itself, by relying on information technologies, which greatly facilitate the gazing experience. The self uses the digital imaginary concerning itself to construct a virtual identity through which it seeks to grasp its own personal identity (the question "who am I for you?" becomes "who am I online?"), in a potentially feedback loop of adjustments and modifications leading to an equilibrium between the off-line and the online selves. The observing is normally hidden and certainly not advertised. And yet, by its very nature, the digital gaze must be understood both as an instance of presumed "common knowledge" of the observation ("I know that you know that I know etc. … that this is the way I am seen by you") and as a private experience (it is still *my* seeing of myself, even if I try to make sure that such seeing is as much like your seeing as I can). The digital translation of the gaze has important consequences for the development of personal identities. First, there is the amplification, postponement (in

terms of age), and prolongation (in terms of duration) of the gazing experience. This means that the *ontic feedback* – the tendency of the gaze to re-ontologise (change the very nature of) the self that is subject to it becomes a permanent feature of the onlife experience. Second, through the digital gaze, the self sees itself from a third-person perspective through the observation of itself in a proxy constrained by the nature of the medium, which affords only a partial and specific reflection. Third, the more powerful, pervasive and available ICTs are, the more the digital gaze may become mesmerizing: one may be lost in one's own perception of oneself as attributed by others. And finally, the experience of the digital gaze may start from a healthy and wilful exposure/exploration by the self of itself through a medium, but social pressure may force it on selves that are then negatively affected by it, leading them to re-ontologise themselves eteronomously.

6. The Logic of Realisation

We are coming to the end of our exploration, but before drawing a final conclusion one more topic needs to be covered for the sake of completeness. In the previous pages we have quickly looked at the process of progressive detachment (membranes) of the self from the non-self (the world), and at the role played by ICTs in the construction of personal identities. The process itself, however, is also part of the narrative through which we semanticise reality, i.e., through which we make sense of our environment, of ourselves in it, and of our interactions with it. In other words, the process of progressive detachment of the self from the non-self is reconstructed by the self from the self's perspective. The ultimately internal nature of such perspective is inescapable, but it can be made critically explicit, and this is the concluding move we need make.

In order to do so, I suggest we borrow a concept from Aristotle's *Poetics*, that of *anagnorisis*. The Greek word is translated differently depending on the context. In Aristotle, the phenomenon of *anagnorisis* refers to the protagonist's sudden recognition, discovery, or realisation of his or her own or another character's true identity or nature. Through *anagnorisis*, previously unforeseen character information is revealed. Classic narratives in which *anagnorisis* plays a crucial role include *Oedipus Rex*, *MacBeth*, *The Sixth Sense*, *The Others*, or *Shutter Island*. I shall not

spoil the last three, if the reader has not watched them. Generalising, one may say that, given an information flow, *anagnorisis* is the information process (epistemic change) through which a later stage in the information flow (the acquisition of new information) forces the correct reinterpretation of the whole information flow (all information previously and subsequently received). For this reason, I prefer to translate *anagnorisis* as realisation. Figure 1 provides an illustration.

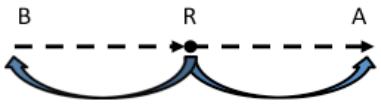

Figure 1. The Logic of Realisation

The logic of realisation should not be confused with the logic of falsification. At point R, some information becomes available that does not make some information at point B (for before) false, but rather provides the right perspective from which to interpret it. For example, at R it is still true that at B x loves y, but now (at R) x realises that it is fraternal love, and this is not going to change in A (for after). The difference should be clear once we see that the information at point R also affects information at point A. Thus, realisation is a concept that belongs more to hermeneutics than to epistemology.

If we now apply the logic of realisation to the development of the 3C model, we may understand that it is the self that is speaking about itself, and then appreciate that it is actually through the self that information becomes self-aware. Let me be less abstract. In a different context (Floridi 2008b, 2011b), I have defended a view of the world as the totality of informational structures dynamically interacting with each other. If this is the case – or at least in order for a philosophy of personal identity to be consistent with such a view – selves too must be interpreted as informational structures. Selves are the ultimate negentropic technologies, through which information temporarily overcomes its own entropy, becomes conscious, and able to recount the story of its own emergence in terms of a progressive detachment from external reality. There are still only informational structures. But some are things, some are organisms, and some are minds, intelligent and self-aware beings. Only

minds are able to interpret other informational structures as things or organisms or selves. And this is part of their special position in the universe.

7. Conclusion: from the Egology to the Ecology of the Self

ICTs have made possible unprecedented phenomena in the construction of the self. Self-poiesis today means tinkering with the self, with still unknown and largely unassessed risks and rewards. Amazing as all this already is, we are witnessing only the beginning of an information revolution, which may have even more radical consequences in our self-understanding and the constructions of our own identities. It is, as they say, an interesting time in which to live. In the previous pages, I have outlined what may be a fruitful approach to start understanding the construction of personal identities in onlife environments. Who we are and can be in the infosphere is a complicated and challenging issue, and I am fully aware that much more can and should be done in order to develop our new egology. More philosophical insight and better understanding are needed in order to cope successfully and fruitfully with the new affordances, constraints, and challenges brought about by the exponential development of digital technologies. Unfortunately, as if this were not already a gigantic task, it needs to be paralleled by the development of an equally robust ethics of self-poiesis, a new ecology of the self fully adequate to meet the demands of a healthy life spent in the infosphere. There is much that needs to be done on the ethical front as well. All this won't be easy, but it can be done, and it is certainly worth a try.

Acknowledgements

The research for this article was funded by an AHRC grant on "The Construction of Personal Identities Online". Previous versions of it were discussed at the following meetings: "Who am I Online?", 10-11 May 2010, University of Aarhus, Kaløvig Centre, Denmark; "Personal Identities Online and Information Ethics", 21-22 May 2010, Department of Philosophy, Bilkent University, Ankara, Turkey; "Identity in the Information Society, Third Workshop", 26-28 May, 2010, Hotel Victoria, Rome, Italy; E-CAP 2010 Conference, 4-6 October 2010, Technical University Munich, Germany; Research seminar, 7 December, 2010 Philosophy Department, University of Reading (Reading); Research seminar, 11 March 2010, Balliol College, Oxford, UK; "Personal Identities after the Fourth Revolution", 17 June, 2011, University of Hertfordshire, UK. I am grateful to the organisers and the participants for their feedback and the opportunity to improve the ideas presented in this article. I am sure I should have taken better advantage of it. A final thanks goes to Gregory Wheeler for his feedback on the last version of this paper.

References

Bond, Alan H., and Les Gasser, eds. 1988. *Readings in distributed artificial intelligence*. San Mateo, California: Morgan Kaufmann.

Churchland, Paul M. 1999. Densmore and Dennett on Virtul Machines and Consciousness. *Philosophy and Phenomenological Research* 59 (3):763-767.

Densmore, Shannon, and Daniel Dennett. 1999. The Virtues of Virtual Machines. *Philosophy and Phenomenological Research* 59 (3):747-761.

Floridi, Luciano. 1995. Internet: which future for organized knowledge, Frankenstein or Pygmalion? *International Journal of Human-Computer Studies* 43:261-274.

———. 2005a. Consciousness, Agents and the Knowledge Game. *Minds and Machines* 15 (3-4):415-444.

———. 2005b. The Ontological Interpretation of Informational Privacy. *Ethics and Information Technology* 7 (4):185 - 200.

———. 2006. Four Challenges for a Theory of Informational Privacy. *Ethics and Information Technology* 8 (3):109-119.

———. 2007. A Look into the Future Impact of ICT on Our Lives. *Information Society* 23 (1):59-64.

———. 2008a. Artificial Intelligence's New Frontier: Artificial Companions and the Fourth Revolution. *Metaphilosophy* 39 (4/5):651-655.

———. 2008b. A Defence of Informational Structural Realism. *Synthese* 161 (2):219-253.

———. 2008c. The method of levels of abstraction. *Minds and Machines* 18 (3):303-329.

———. 2010. *Information - A Very Short Introduction*. Oxford: Oxford University Press.

———. 2011a. The Fourth Technological Revolution. TEDxMaastricht, http://www.youtube.com/watch?v=c-kJsyU8tgI&feature=autofb.

———. 2011b. *The Philosophy of Information*. Oxford: Oxford University Press.

Hayes, Patrick, Stevan Harnad, Donald Perlis, and Ned Block. 1992. Virtual Symposium on Virtual Mind. *Minds and Machines* 2 (3):217-238.

Hume, David. 2007. *A treatise of human nature : a critical edition*. Vol. 1-2. Oxford: Clarendon.

Locke, John. 1979. *An essay concerning human understanding*, *The Clarendon edition of the works of John Locke*. Oxford, New York: Clarendon Press; Oxford University Press.

Martin, Raymond, and John Barresi. 2006. *The rise and fall of soul and self : an intellectual history of personal identity*. New York, N.Y.: Columbia University Press.

Minsky, Marvin Lee. 1986. *The society of mind*. New York: Simon and Schuster.

Perry, John. 2008. *Personal identity*. 2nd ed. Berkeley, CA; London: University of California Press.

Pollock, John. 2007. What am I? Virtual Machines and the Mind/Body Problem. *Philosophy and Phenomenological Research* 76 (2):237–309.

Proust, Marcel. 1992. *In search of lost time. 1, Swann's way*. London: Vintage, 1996.

Putnam, Hilary. 1960. Minds and Machines. In *Dimensions of Mind*, edited by S. Hook. New York: New York University Press.

Schechtman, Marya. 1996. *The constitution of selves*. Ithaca ; London: Cornell University Press.

Sloman, Aaron, and Ronald L. Chrisley. 2003. Virtual machines and consciousness. *Journal of Consciousness Studies* 10 (4-5):133-172.
Sorabji, Richard. 2006. *Self : ancient and modern insights about individuality, life, and death*. Chicago: University of Chicago Press.
Sycara, Katia P. 1998. Multiagent Systems. *AI Magazine* 19 (2):79-92.
Turkle, Sherry. 1995. *Life on the screen: identity in the age of the Internet*. New York: Simon & Schuster.
Warren, Samuel, and Louis D. Brandeis. 1890. The Right to Privacy. *Harvard Law Review* 193 (4).
Wooldridge, Michael J. 2009. *An introduction to multiagent systems*. 2nd ed. Hoboken, N.J.: Wiley; Chichester.