# Identity and Self-Knowledge

John Perry[*]

*Self*, *person*, and *identity* are among the concepts most central to the way humans think about themselves and others. It is often natural in biology to use such concepts; it seems sensible to say, for example, that the job of the immune system is to attack the non-self, but sometimes it attacks the self. But does it make sense to borrow these concepts? Don't they only pertain to *persons*, beings with sophisticated minds, and perhaps even souls? I argue that if we focus on the every-day concepts of self and identity, and set aside loftier concepts found in religion, philosophy, and psychology that are applicable, at most, to humans, we can see that *self* and *identity* can be sensibly applied widely in biology.

**Keywords**

self • self-knowledge • identity • person • personal identity • unity-relation • process philosophy

*Part of a special issue, Ontologies of Living Beings, guest-edited by A. M. Ferner and Thomas Pradeu*

**Editorial introduction:** John Perry works primarily in logic, metaphysics and the philosophy of mind—and this essay is a contribution to a sub-field in the latter: *self-knowledge*. How do we have, and how do we understand, knowledge of ourselves and of our own mental states? While initially this might seem quite far removed from the focus of this special issue, there are two important reasons for including his piece here.

Firstly, the essay sets out and examines different uses of the term 'self' and 'identity' (alongside the related notion of 'genidentity'). Given the aims stated in the general introduction, the utility of this should be clear; these terms are used in a variety of ways, and Perry's analysis provides a clear introduction, which helps us guard against confusions borne from polyvalent expressions. (In doing this it stands alongside Christopher Shields's piece, 'What Organisms Once Were and Might Yet Be,' which maps out the different nuances of the 'organism' concept—and we hope such papers as these will invite philosophers of biology (and *biologists*) to offer similar definitions of similar terms, e.g. 'biological.')

Secondly, and perhaps more importantly, Perry organises these definitions in relation to discussions on going in the philosophy of biology (and biology more generally). Specifically,

[*]Cordura 127, CSLI, Stanford University, 210 Panama St, Stanford, CA 94305-4101, USA,
john@csli.stanford.edu

Received 1 October 2016; Accepted 22 October 2016
doi:10.3998/ptb.6959004.0009.005

he focuses on how we might best interpret the so-called 'self/no-self' model of immunology (of the kind explored by Pradeu and Carosella), but his analysis of the notion of 'self' is also, of course, applicable to discussions of *self-renewal* (of the kind found in Melinda Fagan's examination of stem cell lineages—'Stem Cell Lineages: Between Stem and Organism'—also in this issue). In addition, Perry's essay provides a useful partner piece to Rory Madden's 'Animal Self-Awareness.' Both examine a characteristic question of the 'personal identity' debate: *What am I*? We are aware of ourselves, but our self-knowledge has limits—and the character of the self to which we refer when we say 'I' is not immediately obvious. (Are we *purely psychological beings*? Are we *animals*? Might we be both?) It is at this juncture that Perry's contribution begins to bleed into the focus of the special issue proper. It offers insight into our ways of referring to things in the world while simultaneously emphasizing how, as knowers, we are often and easily confused. In pursuing metaphysical inquiry, in proposing ontologies of living beings, we must be aware of these pitfalls in order to avoid them. –AF/TP

> The self/non-self model is no longer an appropriate explanation of experimental data in immunology … this inadequacy may be rooted in an excessively strong metaphysical conception of biological identity. We suggest that another hypothesis, one based on the notion of continuity, gives a better account of immune phenomena. Finally, we underscore the mapping between this metaphysical deflation from self to continuity in immunology and the philosophical debate between substantialism and empiricism about identity.
>
> –Pradeu and Carosella (2006)

## 1 Introduction

The word 'self' seems rather schizoid. It has a lofty use. In a philosophical or religious text, the self may be taken to be the the deepest aspect of human existence: the mind, or the soul, or a key aspect of both. In psychology, one's self may be those properties and traits one or believes oneself to have, that are most important and essential to who and what one is. Self-knowledge, in either construal of the self, is important and perhaps hard to achieve. It's not something we expect of inanimate objects, like cars, or even living systems of a kind we humans like to regard as more primitive than ourselves, like wisteria bushes, or magnetosomes, or chickens. And whatever the right account of the immune system, it doesn't seem that recognition of one's mind, or soul, as self rather than non-self could have much to do with it.

But the English morpheme 'self' in its most common uses, seems to have a much more modest job. It is not quite a word, but something that makes an ordinary pronoun into a reflexive one: 'her' into 'herself,' 'him' into 'himself,' and 'it' into 'itself.' If I say my car is destroying it*self*, because of the vibration caused by its unaligned wheels, I am not referring to some part or aspect of the car, its *self*, that thinks and acts and might end up in heaven, nor to some aspects of my car that my car finds very important. I am just referring to my car, as the thing being destroyed, in a way that makes it clear that it is also the thing causing the destruction. If I say the wisteria on my trellis is falling all over it*self*, I am just making it clear that the wisteria which falls and the wisteria which it falls on are parts of the same plant.

When immunologists debate whether the job of the immune system is to protect self from non-self, or talk about self-organizing systems, is it a lofty or a mundane sense of 'self' that is being put to use? Or is it perhaps a misleading use of the word which imports, using ambiguity veiled by unwitting metaphor, explanations and descriptions appropriate for humans into discussions of much different forms of life? I think it is rather an appreciation of more humble uses of 'self,' which are applicable to all aspects of life, and central to understanding all of its uses, even the most lofty.

In this essay I do not claim to solve, explain, or even understand, the various issues and debates in biology in which the expression 'self' may be used. I do hope to illuminate the connection between the lofty and not so lofty uses of 'self,' and among various sorts of 'self-knowledge,' in a way that might be relevant to such issues and helpful to those who discuss them.

In particular, I consider two problems that have led philosophers to metaphysical worries about selves, and complicated theories to assuage those worries: personal identity and self-knowledge. I argue that their solution lies instead in the appreciation of the more humble, and widely applicable, uses of 'self' and 'identity.'

## 2 Self, Identity, and Unity

Reflexive pronouns such as 'itself,' 'himself,' and 'herself' are used when the object of an action or attitude is the same as the subject of that action or attitude. If I say Mark Twain shot *himself* in the foot, I describe Mark Twain not only as the shooter but as the person shot; if I say Mark Twain admired *himself*, I describe him not only as the admirer but as the one admired. In this sense, 'the self' is just the person, or creature, or thing, that is both the subject and the object of the attitude or action in question. Mark Twain, as subject, doubtless admired many people over the years. Mark Twain, as object, continues to have many admirers. The pronoun 'himself' allows us to get at the case where he is both subject and object of admiration. 'Self' is also used as a prefix for names of activities and attitudes, identifying the special case where the object is the same as the agent: self-love, self-hatred, self-abuse, self-promotion, self-knowledge, and suicide.

When we say 'the same' and 'identity' in these contexts, we mean that there is *only one thing*. The way I shall I use '$A$ and $B$ are identical,' it means there is just one thing that both *is* $A$ and *is* $B$. Mark Twain and Samuel Clemens are identical because there is just one fellow that was Mark Twain and was Samuel Clemens; one fellow, two names. Both words, 'identity' and 'same,' are also used for various forms of similarity. When we say that twins are *identical*, we do *not* mean that there is just one of them. If you are washing your car, and I say, 'I'm going to do the same thing,' I probably mean I am going to wash *my* car, not that I am going to help you wash yours. I'm going to do something similar to what you are doing, but there will be two similar car-washings, not just one.

Some philosophers, notably Heraclitus around 500 BCE, argued that that similarity is the most we can find in the world; identity rules out change, but change is everywhere, so identity is nowhere. 'You cannot step in the same river twice, because new waters are rushing in' (Heraclitus 2003). Heraclitus held a view that now might be termed 'process philosophy.' The world consists of events, happening one after the other; change and flux are the rule; permanent objects are illusory; explanation always rests on relations among events, not on the dubious category of *things*.

Modern process philosophers, like Whitehead, agree about the importance of events and the ubiquity of change, but do not see permanent or semi-permanent objects as illusions (Whitehead 1929). Such things are real processes made up of events that bear important relations, especially causal relations, to one another. Our concepts of things are based on these relations and their importance; such processes exhibit significant patterns we can recognize and use to cope with the changing world.

I think one can step in the same river twice, even though not only the river stepped into, but also the person doing the stepping, consist of events and exhibit continual change.

But how can the same river contain, and even consist in, different water at different times? A Heraclean might appeal to the self-evident principle of the *indiscernibility of identicals*. If $A$ and $B$ are identical—that is, that there is just one thing that is both $A$ and $B$—then *if* $A$ has a certain property, $B$ has it too. After all, if there is just *one thing* that is both $A$ and $B$, that one thing can't both have and not have a given property. Doesn't it follow that if the river I stepped in yesterday is the identical river I step in today, and the river I stepped in yesterday contained, say, water molecule 73, then the river I step in today must also contain water molecule 73? Such arguments have led philosophers to think that if persons are to have *real* identity, there must be unchanging substances or essences involved.

But the argument is no good. Objects have properties *at times.* An object cannot both have and lack a property at a given time. But an object can have a property at one time, and lack that same property at another time. The river I step in today *contained* water molecule 73 *yesterday*, even though it doesn't contain it today. And the river I stepped into yesterday *does not contain* water molecule 73 *today*, even though it contained it yesterday. There is just one river. And similarly, there was just one person, one self, namely me, who stepped in the river yesterday and did so again today. Our concepts of objects and their properties allow an object to change, in the sense of having different properties at different times; this does not conflict with any sound principle of logic or metaphysics.

That does not mean there are no constraints on how an object can change. You can't step in the same river twice by stepping into a North American River Monday and and African river Wednesday. Rivers can change a lot in two days, but not by being in one continent on Monday, and on another continent, separated from the first by an ocean, on Wednesday. And something that is a river on Wednesday will not change into a building, or a human being, or number no matter how long we wait. Where do these constraints come from?

Identity, I said, means that there is just one object. That seems simple enough, but can be puzzling. Is it a relation at all? After all, if it is true that $A$ is identical with $B$, then there is just one thing that is both $A$ and $B$. But don't we have to have at least two objects to have a relation? If we grant that identity is a relation, it seems to be quite universal and trivial: every object is identical to itself and no other. Any number is identical to itself. Any mountain is identical to itself. Any photon is identical to itself. And every person is identical to himself or herself. But deciding issues involved in identity can be a tricky process. Is there really a single trivial relation involved in all of these diverse phenomena?

We need to distinguish between identity and what I will call 'temporal and spatial unity relations.' These unity relations vary with the kind of objects we are considering, and they are where constraints arise. Imagine a simple kitchen table. The table I am imagining has five spatial parts—four legs and a top. These parts are not identical with one another. They are related to one another in various other ways. Any two parts are either attached to another, either directly or indirectly, in such a way as to form a composite object with a certain shape and size, which can fulfill a certain function: it's tall enough for chairs to slip under it, and flat enough and large enough to put plates and glasses on without their slipping off. The ways the parts need to be related at a given time, so that they form a whole with a characteristic shape, size, and function we call a 'table,' constitute the *spatial unity relation* for tables, at least for simple tables like I am imagining.

Although the five parts are not identical, there is just one table of which they are all parts. If I say, pointing in turn to different legs, 'the table of which this is a part, is identical with the table of which that is a part,' I make a true identity statement. At the level of the parts, we have real relations: being attached; being formed from the same kind of wood, etc. But there is just one table, which shares all of its properties with itself, as demanded by the indiscernibility of identicals.

There is also a temporal unity relation for tables.[1] Given our process ontology, this is a relation between events. Consider two events, on successive days, each consisting of a kitchen table occupying a room. If there is just one table, we expect these events to be parts of a spatio-temporally continuous path of such events. At every spatio-temporal point along such a path, there is a table occupying the space at the time, and the successive tables are either very similar, or there is an explanation for the changes—if, say, someone paints the table.

There is an ancient problem about identity, that brings out the importance of the distinction between identity and unity.

> The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same (Plutarch).

In daily life, when we disassemble objects and then reassemble them, we have no problem considering them identical. If the repair shop takes my bicycle completely apart and cleans and reassembles the parts, it's still my bike, the same one I brought in for repair. Even if they replace a few parts, there is no question. It seems that there is a sort of default unity relation for bicycles and other such artifacts. They should be spatially temporally continuous, with their parts intact at each point. But we have some back up relations which take care of identity in the strange cases that can arise with things that can be taken apart and put back together.

In the case of the Ship of Theseus, as Plutarch describes it, the worry is how many parts can the Athenians replace, and still have the same ship? None? Half? All? Perhaps there is no clear answer? And then there is an added worry. What if Athena carefully saved all the planks that were removed and replaced, and then reassembled them into a ship just like the original. Wouldn't her ship, with the original planks, really *be* the original, rather the version honored by Athenians, with all its new planks?

Problems like this have led some philosophers to recommend giving up the concept of identity through time, in place of *genidentity*, a relation between temporal parts. Of course, the temporal parts, if they last any time at all, still have identity, so identity won't go a way completely.

At issue is the transitivity and symmetry of identity. If $A$ is identical with $B$, and $B$ is identical with $C$, then $A$ is identical with $C$. If $A$ is identical with $B$, $B$ is identical with $A$. These properties follow from the fact that we just have one thing all along. Now each part of a

[1] The concept of a temporal unity similar to that of *genidentity*, an idea developed by Kurt Lewin in his *Habilitationsschrift* (Lewin 1922), and used by Reichenbach, Carnap, and others in the philosophy of science. This concept is usually connected with the idea of replacing identity with genidentity for scientific purposes, in part because of problems like the Ship of Theseus, discussed below.

thing, temporal or spatial, provides a way of referring to the thing of which it is a part: 'the ship of which this plank is a part' or 'the baseball game of which this is an inning.' It seems to follow that the unity relations should also be transitive and symmetrical, so that they are equivalence relations (reflexive, transitive, and symmetrical). And usually they are. All the innings played this summer will doubtless fall neatly into games, for example.

But unity relations, unlike identity, are not always *necessarily* equivalence relations. Highways are a good example. Highway 1 goes up the Coast of California. Highway 101 stays inland, sometimes a few miles east of Highway 1, sometimes many miles. They merge to cross the Golden Gate Bridge. If the unity relation just requires a path from one highway part to another, then all the parts of both highway will have it to one another, and we won't have two highways after all, just one, with different names for different parts. This might lead us to posit a more complex relation for highway temporal unity. $A$ and $B$ are parts of the same highway if you can get from one to the other without leaving the road or making a U-turn, perhaps. So we have two highways, that share a part. Then when I am on the Bridge, am I on two highways? It doesn't *seem* like it.

Does it matter much which we of looking at it we choose? Perhaps, because Highway 1 is a state highway, while Highway 101 is a Federal highway. The state of California might argue that the first option is right, there is just one highway, and the Federal government should pay for everything. Legal issues aside, it doesn't seem like a deep problem.

Seeing genidentity as a replacement for identity is motivated by the idea that such indeterminate and puzzling cases cannot be tolerated in science, particularly axiomatic science. But in every day life with every day objects, it's seldom a problem. We all understand, at least implicitly, how the identity of objects rests on unity relations. When the concept of identity doesn't quite fit certain cases, we revert to thinking in terms of parts and unity relations.

In the case of persons, it is easy to come up with *possible cases* in which plausible temporal unity relations for persons turn out not to be equivalence relations (Perry 1972). These possibilities are crises for those who place personal identity in an immutable essence or immaterial soul (see below). On my view, they work to remind us that even with persons, the concept of identity is applicable because the underlying relations of unity are well-behaved. Perhaps, with the march of science, they will become less so.

## 3 Personal Identity

Tables and human beings are both physical things, systems of physical events, related to one another through space and time. But of course humans have many properties that tables don't. We are complex biological systems. We have sentience and intentionality. That is, we perceive and feel, and we have beliefs and desires.

Descartes argued that no physical system can have such properties as sentience and intentionality (Descartes [1641] 1993). We need to see minds as a separate sort of things from physical bodies. These non-physical minds, Descartes thought, were immaterial substances, which could be identified with the souls needed by Christian theology. My mind/soul/self can survive, when my body and all of its parts have turned to dust and ashes. Minds are not made out of dust and ashes in the first place, so they don't need to return to dust and ashes when we die. $A$ and $B$ are the same person, if they have the same mind.

While no philosophical views, as far as I can tell, are ever completely abandoned, 'Cartesian dualism' is not very common these days. However, what is called 'property dualism' is advocated by many, who provide interesting arguments for it. The idea is that such things as being in pain, or seeing a sunset, or believing that Albany is the capital of New York—properties involved in sentience and intentionality—*are* properties a human has in virtue of the states and properties of his brain or central nervous system. But they are not *physical* properties of the brain. I won't worry much about property dualism. For one thing, I've already done the best I can to refute the arguments for it (Perry 2001). For another, the issue is somewhat orthogonal to the main items I wish to discuss.

What is the temporal unity relation for persons? John Locke argued that even if there are immaterial substances of the sort Descartes postulated, they won't serve to define personal identity, as we have no way of knowing, and no reason for caring, whether or not the same immaterial substance 'thinks in us' from time to time, immaterial substances being immaterial, invisible, and pretty much wholly mysterious (Locke 1690). The view he advocated fits much better with the process perspective and the points I have been making. In my terminology, he thought that the temporal unity relation for persons was *memory*—in particular memories of past thoughts and actions. Memory gives us a relation between a current experience—the remembering—and a past experience, thinking something or doing something. The person who does the remembering is the identical person who did the action, or thought the thought.

Locke's theory has been modified, improved, and defended in many ways in the face of many objections (see Perry 2008). For my purposes, the key point is that it gives us an account of personal identity that does not require the postulation of immaterial substances, interactions between material and immaterial worlds, and other mysteries. In a moment I will argue that it fits into what one might think of as a very biological picture of the self.

## 4 The Self as Subject and Object

One motivation for somewhat metaphysical, and somewhat mysterious, accounts of the self is 'the problem of the self as subject.' Suppose I think that I live in Palo Alto. I play two roles in this episode of thought. I am the *subject*; that is the agent who thinks the thought. And I am the *object*, the thing the thought is about. For various reasons, philosophers have found the self as subject mystifying. When he looked around, Wittgenstein thought he could find himself as object—he could see his legs, for example. But he couldn't find himself as subject: 'The thinking, presenting subject; there is no such thing' (Wittgenstein 1922). Tom Nagel draws a conclusion at least as surprising, that the thinking self is outside of space and time, but sees the world through the perspective of a particular human body (Nagel 1983).

We cannot solve the problem of the self as subject by participating in the lofty and metaphysical discourse about the self. The solution lies in the humbler uses of 'self.'

I distinguish three kinds of knowing (and not knowing) things about oneself, which I call *primitive self-knowledge*, *self-knowledge*, and *knowledge of the person one happens to be*. I claim all organisms have primitive self-knowledge, and it is the basis of self-knowledge.

I begin with an example I've used elsewhere; apologies to anyone who has already encountered it. One day in Vienna, Ernst Mach boarded a bus through the rear door. Looking towards the front, he noticed a rather shabbily dressed, bookish sort of fellow. 'What a shabby pedagogue is that,' he thought to himself. He didn't realize he was seeing his own reflection in a mirror of the sort conductors use to keep track of things when buses are crowded (Mach 1914, 4n).

Assuming his verdict was accurate, at this point Mach had what I call 'knowledge of the person he happened to be.' There is a certain person, the one he saw in the mirror, whom he knew to be a shabby pedagogue. And, unknown to him, that person happened to be Mach

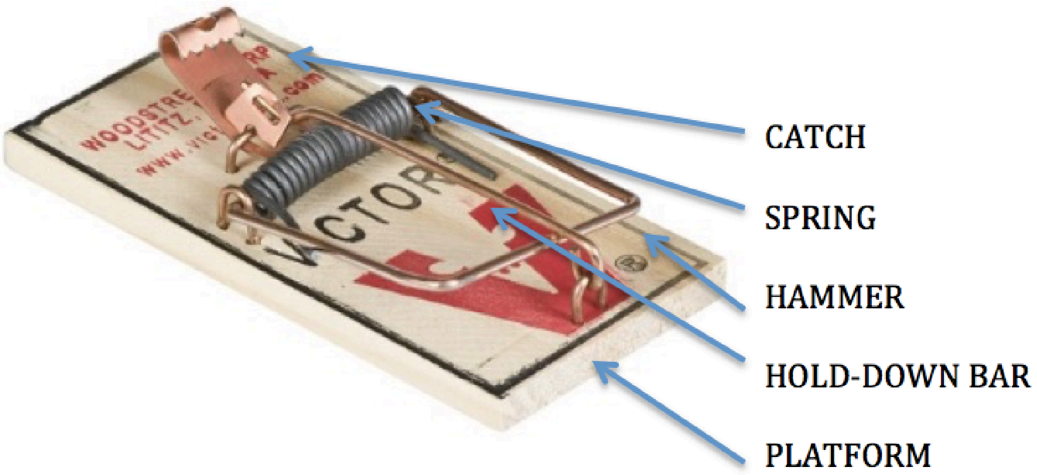

Figure 1: A mousetrap.

himself.

We wouldn't ordinarily call this self-knowledge, any more than we call accidentally killing oneself suicide. Still, killing oneself is a necessary part of suicide, and knowing a fact about the person you happen to be is a necessary part of self-knowledge. Assuming that, up until this time, Mach didn't think of himself as a shabby pedagogue, it was more a case of self-ignorance. Mach didn't believe that *he* was a shabby pedagogue, because he didn't believe that *he* was the person he was looking at. The italicized 'he' warns us that we have a certain way of believing things about oneself in mind: the sort of case where a person would use the word 'I' to express the belief.

After a bit, Mach realized what was going on, that *he* was seeing *himself* in the mirror. Then he thought to himself, 'I am a shabby pedagogue!' This is what I call self-knowledge.

So what happened when Mach went from simply knowing, about a certain person who happened to be him, that that person was a shabby pedagogue, to having what we would ordinarily call *self*-knowledge, that is, knowing that *he* was a shabby pedagogue—having the sort of knowledge he would express with the first person: 'I am a shabby pedagogue'? The sort of knowledge that has Mach as both subject and object?

The difference, I want to claim, is that Mach incorporates the knowledge that a certain person is a shabby pedagogue into the file of things that he knows about himself *primitively*. To explain this, I need to introduce the concept of *harnessing information*. And to do that, I am going to consider an artifact: an ordinary mousetrap. A rather macabre example, I fear; perhaps a case where death helps us appreciate life.

I'll remind you how mousetraps like this one work. One puts some cheese on the catch. A hold-down bar attached to the back of the platform is placed over the hammer and then secured under the catch. When a mouse nibbles on the cheese, the catch moves, the bar is released, the hammer is freed, and due to the spring, smashes down on the front of the trap, killing the mouse.

Here is a way of looking at this as harnessing information. First, information. Consider an event, $x$'s being $\phi$ at $t$. This may carry the information that another event, say $y$'s being $\psi$ at $t'$ has occurred, is occurring, or will occur. That is, given the way the world works, at least for a time in a region—the constraints—and the particular circumstances in the $t$–$t'$ interval, an event of the first type significantly increases the probability of an event of the second type.

What counts as 'significant' depends on what we are talking about.

With the mousetrap, the constraints involve the strength of springs, the nature of mouse bodies, the width of the wire of which the hammer is made, the behavior of mice, as well as such things as size of cats and human feet. In the case of mousetraps the user chooses the circumstances, and so, to a certain extent, the constraints. Suppose I place the trap in the narrow space between my refrigerator and the wall. I place it back far enough that my toe won't hit the catch when I get something from the refrigerator. The space is narrow enough that the cat can't get in far enough to stretch a paw the cheese. Given the way mice work, and these circumstances and others, the catch moving will often carry the information that there is a mouse at the front end of the trap.

How much does the catch moving have to increase the probability of there being a mouse in a position to be killed, in order to count as 'significant'? Not much; for ordinary mouse traps the cost of false positives is not very high. If the traps were expensive, and could only be used once, that would change things.

Now consider the purpose of the trap. I am using it to catch and kill mice. That's what's supposed to happen when the hammer is released. We can distinguish between purpose and the conditions of success. The hammer smashing down on the front of the trap will succeed in its purpose only under certain conditions, namely, if there is a mouse situated there.

So the movement of the catch, by carrying the information that there is a mouse at the catch, carries the information (more or less) that the conditions of success for the snapping of the hammer are satisfied. The architecture of the trap *harnesses* this information. The size of platform, the shape and size of the hammer, the force of the spring, and the length of the hold-down bar ensure (more or less) that the hammer will be released when and only when there is a mouse in a position to be killed. Or at least, these factors make it reasonably probable that if there is a mouse there the hammer will be released, and if the hammer is released it is reasonably probable that there is a mouse there.

I don't want to say that the mousetrap has primitive self-knowledge. Primitive self-knowledge is what we get when Nature harnesses information; that is, when the processes of evolution result in a system that (a) has a repertoire of actions that can promote a natural value, such as survival or reproduction, and (b) has the capacity to pick up information about the circumstances of success of this action, and (c) has an architecture that harnesses that information.

Values require a valuer, it seems. For these cases, we invent one: Mother Nature. But that's a metaphor. And so is my phrase 'natural value.' To think of naturally evolving systems as harnessing information, we need to find four things: information that is detected, actions that are caused, an architecture connecting these, and an 'end-state' that is promoted by this arrangement. It's nice to think of these end-states chosen by Mother Nature, in accordance with her values, although if we pursue the metaphor too far, Mother Nature will turn out to be rather an odd duck.

A simple example of such natural harnessing is the magnetosome (Dretske 1986). Built-in magnets detect which direction is magnetic north; the architecture of a magnetosome determines that it swims in that direction. In the northern hemisphere, such swimming takes the magnetosome into deeper water, where there is less oxygen, which is the situation the magnetosome needs to survive.

Or consider a chicken, looking at a kernel of grain in the barnyard. The state of the chicken's eyes carry a lot of information, in the sense that the eyes wouldn't be in that state, or at least it would be very improbable that they would be, unless certain other things are true. Their being in that state may carry the information that life has evolved, that there is a sun, an so on. But, following Dretske (1986), what the chicken *perceives* is that information that is *harnessed* by the chicken's architecture.

The chicken perceives the presence of an edible kernel. This is the information carried by the state of its eyes that is harnessed. In a well-run barnyard, at least, where there won't be any kernels toxic to chickens lying around, the chicken's eyes wouldn't be in that state unless there were something nourishing in front of it, waiting to be pecked. If the chicken is hungry, which in my experience chickens almost always are, being in this state will lead it in a way that leads to digestion of the nutritious corn or millet. The architecture of chickens harnesses information for the end of survival of the chicken in question.

The chicken has what I call primitive self-knowledge. The information that the chicken detects is information about *itself*—using self in a perfectly modest way, and not as a hidden metaphor for selves or souls or complex minds of the sorts that humans have that allow them to admire themselves, hate themselves, and have identity crises. It is information about the the direction and distance of the kernel of corn from the very chicken who sees it that is harnessed. And it is harnessed for the benefit of that very chicken—it is the one which gets nourished as a result of pecking in that situation. So the chicken does have information about itself. It has *primitive* self-knowledge. This does not require mastery of the first-person, or some special inner version of 'I.' It just requires an architecture that allows pick up of information about the chicken, and harnesses that information to cause actions by that chicken, that promote the relevant values.

When the chicken perceives a kernel of corn, it perceives things concerning itself: that there is a kernel of corn a certain distance and direction from *it.* Any chicken could be in the same state, but by being in that state, each chicken has information about itself. I call such states *normally self-informative.*

Similarly, when the chicken pecks at the kernel of corn in front of it, the effects of the pecking, if conditions for success are present, are effects on that same chicken: it ingests a kernel of corn and it gets nourished. This is a *normally self-effecting* way of acting.

Chickens are among the animals that do not pass the 'mirror test.' A chicken looking at a mirror is perceiving itself. But it is not doing it in a normally self-informative way, but rather via states that would normally be caused by seeing other chickens which were in close in front of it, and looking right at it. And it acts accordingly, with aggressive behavior. Great apes, including chimpanzees and humans, pass the mirror test (Gallup 1970). If you put a chimpanzee in a room with a full length mirror, it will initially exhibit some aggression, like the chicken. But with a little time, other behaviors emerge. In particular, if you anesthetize the chimpanzee, put some colored stuff above its eyebrow and behind its ear, when it awakes, and sees its reflection, it will engage in normally self-effecting action, using its own arms to rub the smudges, a technique that wouldn't work to clean up other chimps.

What a chimpanzee can do, and a chickens can't, is to *integrate* information about itself obtained in ways that are not normally self-informative—information about the chimpanzee it happens to be—with information obtained in self-informative ways. Thus information of the former kind motivates normally self-effecting actions, as in the case of the smudged chimpanzee.

Now let us return to Mach. Mach doesn't pass the mirror test at the beginning of the episode.

Suppose the shabbiness in question includes having a lot of lint on one's vest. When Mach sees lint on his own vest by looking downwards at the front of his clothed body, he will more-or-less automatically swipe it off with his hand. Normally self-informative way of perceiving lint-on-one's-vest, normally self-effecting way of removing it. At the beginning of the episode, Mach picks up information about the lint on Mach's vest, in a way that is not normally self-informative. But his information is not integrated with his self-knowledge. He does not say 'I have lint on my vest,' a normally self-referring way of speaking. And he does not wipe the lint of his vest with a normally self-effecting swipe of the hand.

But then, in a bit, he does pass the test. Like most of us, Mach had a pretty good idea of what he looks like to others, so perhaps he first recognized himself, and then figured out that he must be looking in a mirror. Or perhaps he noted an agreement between the movements he initiated and the ones the fellow he saw made. Or perhaps he just remembered that busses have such mirrors, and went from there. In any case, he integrates the information he perceived about the man in the mirror with the information he has about himself via normally self-informative ways, and acts accordingly. The difference between merely having knowledge about the person one happens to be, and having self-knowledge, is primitive self-knowledge. To recognize that the person one has learned about is one*self*, the person we each call 'I,' and normally worry about in a rather special way, is simply to integrate the information gained on this occasion with the information gained in normally self-informative ways.

## 5 Conclusion

Understanding the cluster of philosophical problems around persons and personal identity, selves and self-knowledge, depends on eschewing the metaphysical constructions, and seeing the concepts as reflecting very basic structures and processes, of biological phenomena. There is no doubt that biology can be helpful to philosophy. The naturalness of using such concepts as 'self' and 'non-self' in biology suggests that clarity on these matters might be helpful, so philosophy can return the favor.

## Literature cited

Descartes, René. (1641) 1993. *Meditations on First Philosophy*. Translated by Donald A. Cress. Cambridge: Hackett Publishing.

Dretske, Fred. 1986. "Misrepresentation." In *Belief: Form, Content, and Function*, edited by R. Bogdan, 17–36. Oxford University Press.

Gallup, Gordon G., Jr. 1970. "Chimpanzees: Self-Recognition." *Science* 167: 86–87.

Heraclitus. (500 BCE) 2003. *Fragments*. Translated by Brooks Haxton. Viking Penguin.

Lewin, K. 1922. *Der Begriff der Genese in Physik, Biologie und Entwicklungsgeschichte*. (Lewin's *Habilitationsschrift*.)

Locke, John. 1690. *An Essay Concerning Human Understanding*, Book II.

Mach, Ernst. 1914. *The Analysis of Sensations*. Translated by C.M. Williams and Sydney Waterlow. Chicago & London: Open Court.

Nagel, Thomas. 1983. "The Objective Self." In *Knowledge and Mind*, edited by Carl Ginet and Sydney Shoemaker, 211–232. Oxford University Press.

Perry, John. 1972. "Can the Self Divide?" *Journal of Philosophy* 69: 463–88. Reprinted in John Perry. 2000. *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press.

Perry, John. 1999. *Problems d'Indexicalité*. Selected essays translated by J. Dokic and F. Preisig. Stanford and Paris: Editions CSLI.

Perry, John. 2001. *Knowledge, Possibility and Consciousness*. Cambridge: MIT Press.

Perry, John. 2008. *Personal Identity*. 2nd enlarged edition. Berkeley: University of California Press.

Plutarch. (75 CE) 1906. "Theseus." in *Parallel Lives*. Translated by John Dryden.

Pradeu, Thomas, and Edgardo D. Carosella. 2006. "The Self Model and the Conception of Biological Identity in Immunology." *Biology and Philosophy* 21: 235–252.

Whitehead, Alfred North. 1929. *Process and Reality: An Essay in Cosmology*. New York: Macmillan.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.

ISSN 2475-3025