

# Can Subjectivity be Naturalized?

*Lynne Rudder Baker*

University of Massachusetts Amherst

The term ‘naturalism’ is often associated with natural science. Here, I shall use it more broadly to apply to the nature of reality: If naturalism is true, then all aspects of reality are “centerless” or nonperspectival.<sup>1</sup> To naturalize some aspect of reality is to show that it is eliminable, or that it is reducible to nonperspectival reality. I am using the term ‘naturalism’ ontologically, to pertain to features of reality, as opposed to, say, ways of knowing or explaining. Naturalism is true if and only if reality is wholly nonperspectival. Naturalism requires that reality can be completely described without use of any first-person terms.

So, I intend the question—Can Subjectivity Be Naturalized?—to be ontological, to raise the question of whether there is a single “centerless” reality or whether there are aspects of reality that entail that there are perspectives.<sup>2</sup> If subjectivity can be naturalized, then all the facts of the world are simply “just there”, to be apprehended impersonally or externally. If subjectivity cannot be naturalized, then there are some facts about the world that are irreducibly subjective, apprehensible only from certain points of view.<sup>3</sup> To put the point a different way, if subjectivity can be naturalized, then there are no irreducible first-person properties; but if subjectivity cannot be naturalized, then the apparent first-person properties really are exemplified as such—i.e., they are irreducible.

My aim here is to show that subjectivity—in particular, what I call ‘the robust first-person perspective’—cannot be naturalized.<sup>4</sup> To do this, I shall first pin down exactly what I am and am not arguing for, by considering a view of John Searle’s.

## What is at Issue

My interest is in reality, not in explanation. John Searle takes consciousness—«inner, qualitative, subjective states and processes of sentience or awareness» that distinguish waking life from dreamless sleep—to be a neurobiological problem.<sup>5</sup> Conscious states are «essentially subjective.» (I agree!) But to think that the essential subjectivity of conscious states puts them out of the reach of objective science, says Searle, is to conflate two senses of the objective/subjective distinction. One sense is epistemic; the other is ontological: Science is objective in the epistemic sense in that its claims can be settled in a way that is «independent of the feelings and attitudes of the investigators». In the epistemic sense, what is subjective cannot be settled without reference to feelings and attitudes.

The other sense of the objective/subjective distinction is ontological; it concerns

---

<sup>1</sup>NAGEL 1986.

<sup>2</sup>NAGEL 1979.

<sup>3</sup>This way of putting the question has roots in Thomas Nagel’s “Subjective and Objective”, NAGEL 1979.

<sup>4</sup>For a more detailed argument, see my BAKER 2013.

<sup>5</sup>SEARLE 2000, p. 559.

existence rather than knowledge or explanation. Conscious items such as tickles and itches have a «subjective mode of existence», in the sense that «they exist only as experienced by a conscious subject.» Entities in the environment, such as mountains and molecules have an «objective mode of existence, in the sense that their existence does not depend on any consciousness.»<sup>6</sup> Indeed, they are objective, not only in the ontological sense, but also in the epistemic sense.

Science, according to Searle, is objective in the epistemic sense, but it may study subjective phenomena. Reality, on Searle's view, has ineliminable subjective aspects, but this ontological subjectivity may be studied objectively. Indeed, these subjective items (tickles and so on) are just biological phenomena, theoretically explainable by biology. Searle notes that «the scientific requirement of epistemic objectivity does not preclude ontological subjectivity as a domain of investigation.»<sup>7</sup>

Since the realm of consciousness, on Searle's view, is the only subjective domain, it seems a little awkward to hold that consciousness "is a biological phenomenon like any other," and hence can be investigated neurobiologically. To Searle, the so-called "hard problem of consciousness"<sup>8</sup> is no problem at all. Skeptics doubt that anyone can ever explain why red things look red or why warm things feel warm. Searle's response is, «We know it happens. We know that brain processes cause all of our inner qualitative, subjective thoughts and feelings.» What's left is the neurobiological question, *How* do brain processes cause subjective states of thought and feeling, and «how, exactly, is consciousness realized in the brain?»<sup>9</sup> How, indeed? If conscious states are essentially subjective, they seem unique. In that case, the neurobiological questions about consciousness hardly seem on a par with other biological questions, like How do microorganisms cause disease symptoms?

Searle argues that because, as he thinks, natural science can explain consciousness, his naturalism is intact. Perhaps his epistemic naturalism is intact, but ontologically, he is no naturalist. Indeed, he explicitly appeals to «ontological subjectivity.»<sup>10</sup> If ontological naturalism were true, then 'ontological subjectivity' would be an oxymoron: Subjectivity would have to be eliminated or cashed out in third-personal terms.

Moreover, an *epistemically* objective science of consciousness, even if successful, would provide no reason to include consciousness in *ontology*. An epistemically objective science of consciousness or of anything else may be an "error theory." An eliminativist about consciousness may well hold that the verdict of an epistemically objective science of consciousness is that consciousness does not exist. (For example, Patricia Churchland observed that «just as it turned out that there was no such thing as impetus, there may be no such thing as awareness.»)<sup>11</sup>

Searle seems to combine ontological subjectivity (there is a "subjective mode of existence") with epistemic naturalism (science can explain everything there is). I do not believe that that combination is viable. Here's why: Assume that if science explains something that exists, then science must countenance its existence. If there is a "subjective mode of existence" and science explains it, then science must countenance a subjective mode of existence. If science countenances a subjective mode of existence, then science is not wholly impersonal. But science is wholly impersonal. So, either science does not explain the subjective mode of existence, in which case epistemic

<sup>6</sup>SEARLE 2000, pp. 561, 563, 564.

<sup>7</sup>SEARLE 2000, p. 564.

<sup>8</sup>Cf. CHALMERS 1995.

<sup>9</sup>SEARLE 2000, p. 366.

<sup>10</sup>SEARLE 1992.

<sup>11</sup>CHURCHLAND 1988, p. 309.

naturalism is false, or there is no subjective mode of existence, in which case there is no ontological subjectivity. Either way, ontological subjectivity rules out epistemic naturalism.

Since my interest is ontological, I shall leave epistemic naturalism behind. What is at issue is whether subjectivity can be naturalized. So, let us turn to the question, What is ontological subjectivity?

## 1 What is Subjectivity?

Ontological subjectivity concerns subjective phenomena that are ontologically significant. What it means for subjectivity to be ontologically significant is that subjective phenomena as such must be mentioned in the ontology, a complete inventory of what exists. Concrete entities, properties, and kinds are the sorts of things that may (or may not) be subjective. In a derivative sense, hypotheses are subjective if they entail that there are concrete entities, properties or kinds that are ontologically subjective.

It is natural to align subjectivity with what is mind-dependent: A phenomenon is subjective if and only if it is mind-dependent. I do not think that this characterization is successful, because what is mind-dependent—what is ontologically dependent on there being minded creatures, creatures with intentions—includes far more than just what is subjective. Elsewhere,<sup>12</sup> I have argued against the fundamentality of the mind-independent/mind-dependent distinction. If that distinction marked a deep division in reality, we should draw the line between phenomena whose occurrence entails that there are minded creatures, and phenomena whose occurrence does not entail that there are minded creatures.

In that case, the line between what is mind-dependent and what is mind-independent is in the wrong place. The realm of the mind-dependent would include not only appearances and thoughts, but also all human artworks and artifacts. An artwork could not exist without beings who have intentions (the artist, people in the artworld); a human artifact, e.g., a clock, has an intended function essentially, and could not exist in the absence of beings who have intentions. (A molecule-for-molecule replica of Big Ben that spontaneously coalesced in outer space would not be a clock.) So, the mind-dependent/mind-independent distinction would put artworks and artifacts in the same category as after-images. However, although after-images are subjective, artworks and artifacts are definitely not subjective—even though they could not exist in a world without minds. If we identified the subjective with what is mind-dependent, the subjective would not be a useful category—it would include not only my refrigerator, but also the original of the US Declaration of Independence (that the Founding Fathers actually signed). So, I take mind-dependence to be a hopeless way to demarcate what is subjective.

Rather than trying to find a line to demarcate the subjective, I shall turn to a paradigm case of something that is subjective—the first-person perspective. If the first-person perspective is ineliminable and irreducible to non-first-person properties, then it cannot be naturalized. After explaining what the first-person perspective is, I shall argue that a key first-person ingredient of manifestations of robust first-person perspectives—a self-concept—cannot be reduced to non-first-person terms.

---

<sup>12</sup>BAKER 2007, p. 18–20.

## 2 What is the First-Person Perspective?

The first thing to notice about the first-person perspective is that it is not an entity. It is a property—a dispositional property, with two stages and countless manifestations. The first-person perspective has two stages: rudimentary and robust. Both stages are capacities. The rudimentary stage, which links us persons to the animal kingdom, is the capacity for consciousness and intentionality; the robust stage, which is the seat of our uniqueness, is the language-based capacity for conceiving of ourselves as ourselves in the first person. Since it is the first-person perspective that I think eludes the grasp of naturalism, let me say more about the rudimentary and robust stages of it.

## 3 The Rudimentary Stage of the First-Person Perspective

The rudimentary first-person perspective connects human and nonhuman animals; it connects animals that constitute persons with other animals. A human infant is a person constituted by a human animal. An infant is born with minimal consciousness and intentionality, which are the ingredients of a rudimentary first-person perspective. A person comes into existence when a human organism develops to the point of being able to support a rudimentary first-person perspective. Unlike the organism, the new entity—the person constituted by the organism—has a first-person perspective essentially.

The rudimentary first-person perspective does not depend on linguistic or conceptual abilities. The rudimentary first-person perspective is found in many species, perhaps all mammals, and seems to be subject to gradation. Consciousness and intentionality seem to dawn gradually (from simpler organisms (like bats?)) and the rudimentary first-person perspective seems to become more fine-grained as it runs through many species in the animal kingdom.

Darwinism offers a great unifying thesis that «there is one grand pattern of similarity linking all life.»<sup>13</sup> Considered in terms of genetic or morphological properties or of biological functioning, there is no discontinuity between chimpanzees and human animals. In fact, human animals are biologically more closely related to certain species of chimpanzees than the chimpanzees are related to gorillas and orangutans.<sup>14</sup>

Human infants, along with dogs, cows, horses and other non-language-using mammals, also have rudimentary first-person perspectives. So, Constitutionalism recognizes the continuity between human animals that constitute human infants and higher nonhuman animals that constitute nothing. In this way, the biological continuity of the animal kingdom is unbroken.

But wait! If that is so, then why do I say that we persons are only constituted by animals and not identical to them? For this reason: Although there is no discontinuity in the animal world—no *biological* discontinuity—the evolution of human persons (perhaps by natural selection) does introduce an *ontological* discontinuity, as we shall see.

The ontological discontinuity between persons and animals lies in the fact that a human infant—but not a human animal that does not constitute a person—has a remote capacity to develop a robust first-person perspective. A human infant, a person, has a rudimentary first-person perspective and also a remote capacity to develop a robust first-person perspective. An animal that does not constitute a person may have a rudimentary first-person perspective, but it has no remote capacity to develop a

<sup>13</sup>ELDRIDGE 2000, p. 31.

<sup>14</sup>DENNETT 1995 p. 336. Dennett is discussing Jared Diamond's *The Third Chimpanzee*.

robust first-person perspective. And this remote capacity distinguishes persons from all other beings.

A remote capacity is a second-order capacity to develop a capacity.<sup>15</sup> For example, a healthy human infant has a *remote* capacity to ride a bicycle. She doesn't yet have the capacity to ride, but she does have the capacity to acquire the capacity to ride a bike. When the young child learns to ride a bicycle, she then has an *in-hand* capacity to ride a bicycle; that is, in certain circumstances (when she has a bicycle available and wants to ride), she actually rides a bicycle and manifests her capacity to ride a bicycle. She may never learn to ride a bike, in which case her remote capacity to ride a bike would not issue in an in-hand capacity to ride a bike.

Similarly, even though a remote capacity to develop a robust first-person perspective is an essential property of persons, it may never issue in a robust first-person perspective (if, for example, the person had a case of severe autism). Moreover, a being without a rudimentary first-person perspective (e.g., an embryo) does not even have the remote capacity to develop a robust first-person perspective—any more than a baby born without legs has a remote capacity to ride a bicycle. Only—but not all—beings with rudimentary first-person perspectives have the remote capacity to develop a robust first-person perspective. Nevertheless, an infant person has not only a rudimentary first-person perspective, but also has the remote capacity to develop a robust perspective; otherwise the entity would not be a human person. So, the ontological difference between persons and animals lies in the robust first-person perspective or in the remote capacity to develop one. The property that pre-linguistic persons, but not nonlinguistic animals, have essentially is the remote capacity to develop a robust first-person perspective.

What makes persons unique is that only persons have robust first-person perspectives. (If a dog learned to talk and acquired the capacity to conceive of itself in the first-person, a new kind of entity would come into existence, a canine-person. But the point would still hold: only persons have robust first-person perspectives.)

To sum up: The rudimentary stage of a first-person perspective is a nonconceptual stage that entails consciousness and intentionality; the robust stage is a conceptual stage that entails the peculiar ability to conceive of oneself as oneself in the first person. The rudimentary stage is what ties us persons to the seamless animal kingdom; the robust stage is what makes us unique. Now let's turn to what, exactly, a robust first-person perspective is.

#### 4 The Robust Stage of the First-Person Perspective

A robust first-person perspective is the capacity to conceive of oneself as oneself in the first-person—without need to identify oneself by means of any description, name, or other third-person referring device. Evolutionarily speaking, my guess is that the robust first-person perspective evolved along with natural language. Conclusive evidence of a robust first-person perspective comes from use of complex first-person sentences like, e.g., 'I wonder how I will die' or 'I promise that I will stay with you.'<sup>16</sup> If I wonder how I will die, or I promise that I'll stay with you, then I am thinking of myself as myself; I am not thinking of myself in any third-person way (e.g., not as Lynne Baker, nor as that woman, nor as the only person in the room) at all. Even if I had amnesia and didn't realize that I was Lynne Baker, I could still wonder how I'm

<sup>15</sup>I found the handy distinction between remote and in-hand capacity in PASNAU 2002, p. 115.

<sup>16</sup>Hector-Neri Castañeda developed this idea in several papers. See CASTAÑEDA 1966, and CASTAÑEDA 1967.

going to die. Anything that can wonder how she—she herself—will die ipso facto has a robust first-person perspective and thus is a person. She can understand herself from “within,” so to speak.

In order to have a robust first-person perspective, one must have a concept of oneself as oneself from the first-person—a self-concept. The second occurrence of ‘I’ in ‘I wonder how I am going to die’ expresses a self-concept. A self-concept cannot stand alone; it is a nonqualitative concept that is used in tandem with other concepts.<sup>17</sup> If I promise that I will take care of you, then I manifest a robust first-person perspective by expressing a self-concept; but also I manifest mastery of other concepts like ‘promise’ and ‘taking care’. And it is in learning a natural language that one masters these other common empirical concepts that one deploys with a self-concept.

Descartes was the inventor of first-person philosophy, but my emphasis on a first-person perspective is far from Descartes’ philosophy. Unlike Descartes, I do not believe in any infallibility of our knowledge of our own minds. Moreover, again unlike Descartes, I have no ambition to be presuppositionless nor to aim for absolute certainty: my view has many empirical presuppositions about ourselves, our language and our environment. One acquires a self-concept by learning a natural language. To acquire a self-concept, one must have a battery of ordinary empirical concepts like *hungry, fun, water, mama*—concepts that one learns in learning a language. And one cannot acquire a language in isolation. On my view, if Descartes were really alone in the world (except for an Evil Genius), it would have been impossible for him to entertain the thought that he was sitting in front of a fire in his dressing gown. There would be no way for him to acquire concepts like ‘dressing gown’—much less a self-concept—if he were in social and linguistic isolation. So, a robust first-person perspective does not afford the logical privacy attributed to Descartes.

To sum up my idea of the first-person perspective: whereas a *rudimentary* first-person perspective is shared by persons and certain nonhuman animals, a *robust* first-person perspective—the conceptual ability to think of oneself as oneself in the first-person—is unique to persons. Human persons normally traverse a path from the rudimentary to the robust first-person perspective, from consciousness to self-consciousness.

The two stages of the first-person perspective support different kinds of manifestations of the dispositional property that is the first-person perspective—the dispositional property that is manifested in being conscious and in acting intentionally at the rudimentary stage, and in conceiving of oneself as oneself in the first-person at the robust stage. The robust stage presupposes consciousness and intentionality and adds a self-concept, which one acquires when learning a natural language.

Both stages of the first-person perspective are first-personal—the rudimentary stage by default, so to speak, and the robust stage by conceiving of oneself in the first-person—but the robust stage deepens the first-personal aspect of the perspective. It is only at the robust that one can have an inner life, that one can be part of the content of one’s own thought. And thinking about oneself—one’s goals, one’s regrets, and so on—in the first person is paradigmatically subjective. Although I have doubts about naturalizing consciousness, my main argument will concern the self-concept that marks the robust first-person perspective. I shall argue that there are first-person properties, I\*-properties, and that they are neither reducible to non-first-person terms, nor are eliminable. Hence, they belong in a complete inventory of what there is.

<sup>17</sup>I think that this point suggests that Hume’s famous passage, «When I enter most intimately into what I call myself, I always stumble upon some particular impression» HUME 1968, p. 252, does not imply that a self-concept has no extension, only that a self-concept be deployed except with other concepts

## 5 Why Subjectivity Cannot be Naturalized

Subjectivity concerns properties of conscious experience, where conscious experience always *seems* first-personal. Those with robust first-person perspectives have self-concepts. This is so, because a robust first-person perspective is the capacity to think of oneself as oneself in the first person, and that capacity is exercised by the deployment of a self-concept. If I say, 'I hope that I'm healthy,' the second occurrence of 'I' expresses my self-concept.<sup>18</sup> A self-concept is a nonqualitative concept that manifests the capacity to conceive of oneself as oneself in the first person. If I assert, 'I hope that I'm healthy,' can my self-concept be empty, or fail to refer? Might there be nothing in reality that satisfies that concept—just as there is nothing in reality that satisfies the *phlogiston* concept.

What would it take for my self-concept in the proposition expressed by 'I hope that I'm healthy' not to be empty? It would take my having a capacity to conceive of myself as myself in the first person. Just a quick look around shows regular exercise of my (and others') capacity to conceive of oneself as oneself\* in the first person. For example, the second occurrence of 'I' in 'I did not realize that I was being rude' expresses the speaker's self-concept. Therefore, we have good reason to suppose that there is such capacity. So, we have good reason to think that the self-concept is not empty.

Since the self-concept is the concept of oneself as oneself in the first person, the only way that a self-concept could fail to express a property would be for the capacity to conceive of oneself as oneself in the first person to fail to exist. However, such a capacity is manifested in all aspects of life—from making marriage vows to volunteering for a suicide mission. And the capacity is presupposed in yet other areas from suing a landlord to retrieve one's property to changing one's mind. Indeed, knowing what one is thinking or doing in general presupposes that the capacity in question. Any sentence or thought of the form 'I  $\varphi$  that I am F,' where ' $\varphi$ ' ranges over psychological and linguistic verbs, is a direct manifestation of that capacity, and a self-concept is expressed by the second occurrence of 'I' in such a sentence.

Since the capacity to conceive of oneself as oneself in the first-person undoubtedly exists, the concept that makes it possible to manifest that capacity expresses a property—a dispositional property. (It's dispositional because you have that property even when you are fast asleep and not manifesting it.) So, a self-concept expresses a property. The property that a self-concept expresses—an I\*-property—is the capacity to conceive of oneself as oneself in the first-person.

The argument just given is simple. A self-concept is a concept that manifests a certain capacity; an I\*-property is a dispositional property a self-concept expresses.

There is a self-concept.

If there is a self-concept, then there is an I\*-property.

∴ There is an I\*-property.

An I\*-property is manifested every time someone asserts a proposition containing a self-concept. So, there are first-person properties and they are exemplified. Hence, sentences expressing concepts associated with an I\*-property cannot simply be eliminated from a complete description of reality. It remains to show that I\*-properties are not reducible to non-first-person properties.

There are a number of ways to construe reduction. A dominant one today is in

<sup>18</sup>If I have a robust first-person perspective, guaranteed by my use of 'I' in the 'that' clause of a first-person psychological or linguistic sentence, then my uses of 'I' even in simple sentences typically express my self-concept.

terms of functional definition, offered by Jaegwon Kim.<sup>19</sup> Adapting Kim's notion of functional reduction and letting P be the property of conceiving of oneself as oneself in the first person, we have:

(P) Having property P is reducible to having non-first-person properties only if there are non-first-person properties that perform the causal task that P performs.

Having property P is reducible only if (i) there is a particular causal task that having P performs and (ii) there are non-first-person properties that perform that causal task. If the first of these necessary conditions is not satisfied, then neither is the second.

It seems clear that (i) is not satisfied. Property P is a nonqualitative property that seems resistant to characterization in terms of causes and effects. But perhaps a functional-reduction approach could make a different contribution—not in providing an impersonal definition of P but in providing a third-person account of the capacity to have P, the capacity to conceive of oneself as oneself in the first-person.

The comparison here would be with functionalism. Functionalists take belief generally to be definable in terms of functional role, where functional role is determined by the causes and effects of the belief. Although functionalists do not hope to give a complete catalogue of the causes and effects of a particular belief, they do suppose that they can give a higher-level account, an account of the capacity to have beliefs. The capacity to have beliefs is the capacity to have mental representations that combine with other mental representations (desires and other beliefs) to produce action.<sup>20</sup> Perhaps the capacity to conceive of oneself as oneself in the first person can be understood functionally in the third person even if the self-concept itself cannot be understood impersonally. Kornblith and other functionalists have suggested as much.

Kornblith (in conversation) has noted that belief- and desire-like states in a computer may interact and cause behavior, but as Perry argued,<sup>21</sup> only if they contain first-person indexicals. Suppose that a computer program has a special mental representation—'+'—an indexical that is directly connected to action.<sup>22</sup> The symbol '+' plays the role of 'I': e.g., '+ sees Fs within reach' and '+ wants to get Fs' leads, *ceteris paribus*, to '+ reaches for Fs'. The problem, as I see it, is that the special mental symbol, '+', at best could play the I-role, not the embedded I\*-role, the self-concept-role (even if there were such and we knew what they were).

However, a functionalist may reply: Analogously to the 'I' case, there is a different special mental representation for a self-concept—say, '#'. The inferential role of thoughts containing self-concepts, then, may be characterized by the relationship between certain representations containing '#'. For example, my first-person belief that

<sup>19</sup>KIM 2000, p. 116.

<sup>20</sup>I do not think that a functionalist view of belief is adequate. For language users, all conscious beliefs may be linked to verbal behavior: a belief that  $f=ma$ , together with a desire to be cooperative and other beliefs may cause the believer when asked, to say, " $f=ma$ ." However, many (if not most) of our beliefs are not even connected with causing (nonlinguistic) behavior—all beliefs about the past, many about the present ("There are now seven billion people on Earth."), some about the future ("There will be a lunar eclipse in 2015"), and some that are non-temporal (" $f=ma$ "). Even if we are concerned only with first-person beliefs, their action-causing role is matched in importance with their other roles: being vehicles for conveying information ("I was born in Atlanta"), for correcting misunderstanding ("I was here on time, but the door was locked"), getting to know someone or articulating truths about oneself ("I have a fear of flying"), narratives that make sense of our lives ("I had a happy childhood until I met Frank..."), and so on. So, a functional definition of 'belief' (first-personal or not) seems to me not to capture the actual concept at issue. And belief would be just a start: we would also need functional definitions of hope, regret, and so on.

<sup>21</sup>PERRY 1979.

<sup>22</sup>Just for the record, let me note that I am extremely dubious about representations as physical tokens in the brain.



I am a registered voter may be caused by my belief that I received a notice about the polling place. The self-concept can be eliminated by showing the connection between representations in the 'belief box'—between '# received a notice about the polling place' and '# is a registered voter', where tokening of the former caused tokening of the latter. So, it may seem, the inferential role of the self-concept can be characterized by the relationship between '#' representations.

This tack seems inadequate for at least four reasons. (i) Beliefs entailing self-concepts are not particularly connected with action as I-beliefs are said to be. Thoughts containing self-concepts are more typically connected to self-evaluation than to action or perception and hence that their causal connections may not issue in any overt action.<sup>23</sup> Belief is (putatively) functionalized by taking a mental representation in the belief box to be caused by other mental representations and perceptions and actions, and to cause actions. But in the self-concept-case, there are often no observable phenomena causally connected to thoughts containing self-concepts. So, there is already a significant disanalogy between the 'I' cases and the self-concept-cases.

(ii) The story about '#' is not generalizable. Suppose that I hope that I will get a such-and-such job, and that one of the causes of my hope is that I believe that I would make more money in that job; and one of the effects (if I get the job), is that I am glad that I make more money. Now we have the following connected '#' representations: '# will get such-and-such job,' in my hope box, caused by '# would make more money in that job' in my belief box and causing (if I get the job), '# got the job' in my glad box. It is one thing to have inferences among mental representations in one's belief box (if you can swallow the metaphor of a belief box); but it is quite another thing to suppose that there are systematic relations between mental representations in belief-, hope-, glad- and all the other boxes.<sup>24</sup>

(iii) There are too many inferential roles—even too many per person—for thoughts containing self-concepts to be characterized in terms of inferential role of mental representations. Thoughts containing self-concepts are not directly connected to action as simple beliefs and desires are said to be connected to action. The kinds of thoughts containing self-concepts are too disparate for the capacity to have them to be regimented as functionalism requires.

Someone may suggest that each kind of thought containing a self-concept ('glad that I', 'wish that I', 'sorry that I', 'fear that I' and so on) has its own functional role. Not only would the same problem of disparateness arise again—this time for each kind of thought containing a self-concept—but this suggestion would fragment the account of the robust first-person perspective. Such fragmentation would prevent the robust first-person perspective from playing its important phenomenological role in unifying our lives. Moreover, we still would have no reason to accept an unsupported blanket assertion that each kind of thought containing a self-concept has its own functional role—especially in view of the idiosyncratic nature of inferential roles of self-concepts.

(iv) From a functionalist point of view, to conceive of oneself as oneself in the first person is nothing other than to have '#' mental representations that are causally connected with other mental representations. Thoughts containing self-concepts are just the "tokening" of mental representations (with wholesale replacement of certain

<sup>23</sup>BAKER 2011.

<sup>24</sup>Moreover, there is the perennial problem of the semantic interpretation of mental representations. In BAKER 1991, I tried to show that Fodor's view of the semantics of mental representations in terms of 'asymmetric dependence' did not work. Philosophers today have just moved on, as if we either had, or did not need, an account of the semantics of 'mentalese.' I side with those who have given up the notion that we have incommunicable representations in our heads.

occurrences of 'I' by '#' or some other stipulated symbol in mentalese) that causally interact with other mental representations.

So, what it is to think, 'I deeply regret that I misspent the past 20 years,' is just a matter of connections between (in my regret box) '# misspent the last 20 years' and other '#' representations. This puts my deep regret (or my profound love) on the same psychological plane as my momentary preference to go to a movie.<sup>25</sup> This does no justice to the phenomenology or moral significance of thoughts containing self-concepts.

Let me conclude my argument that the self-concept is not reducible to non-first-person properties with a dilemma: Either tokening '#'-mental representations entails the capacity to conceive of oneself as oneself in the first person or not. On the one hand, if tokening of '#'-mental representations does entail the capacity to conceive of oneself as oneself in the first person—if '#' really is just a mentalese translation of a self-concept—then we have not reduced the capacity to conceive of oneself as oneself in the first-person to a non-first-person capacity; we have just translated one first-person idiom into another first-person idiom. On the other hand, if tokening '#'-mental representations does not entail the capacity to conceive of oneself as oneself in the first person, then the capacity attributed by '#'-mental representations does not reduce self-concepts. In neither case do we reduce the capacity to conceive of oneself as oneself\* in the first-person, or the I\*-concept, to non-first-personal terms.

These considerations give us good reason to deny that there is a third-person functional definition of the capacity to have a robust first-person perspective. What is needed for reduction of property P is a third-person characterization of the capacity to conceive of oneself as oneself\* in the first person. I have argued that we have no such third-person characterization. So, I take the argument for the reduction of first-person properties to non-first-person properties to fail.

But perhaps first-person properties like P are eliminable altogether. It follows from my earlier argument from self-concepts to first-person I\*-properties that if I\*-properties are eliminable, then so are self-concepts. But if we didn't have self-concepts, we could not make commonplace distinctions; so self-concepts, and hence first-person properties are ineliminable. To see that we need to distinguish between thinking of someone who happens to be oneself from thinking of someone as oneself in the first-person, consider the following little story:

In the 2012 US Presidential election, a columnist for the *NY Times* nicknamed Mitt Romney "Mittens." Suppose that Romney did not know that, but saw a photo of a man from behind labeled "Mittens". Unbeknownst to Romney, he was the man in the photo. He formed the belief, "I believe that Mittens is the man in the photo," but not the belief "I believe that I (myself) am the man in the photo." This illustrates an important distinction—a distinction that cannot be made without a robust first-person perspective. Hence, a robust first-person perspective is ineliminable without cognitive loss.

If a robust first-person perspective is both irreducible and ineliminable without cognitive loss, it belongs in a complete inventory of reality. So, the dispositional property that is the first-person perspective—and with it subjectivity—are part of the furniture of the world. If that is right, subjectivity cannot be naturalized.<sup>26</sup>

<sup>25</sup>Moreover, this example raises the question of how many boxes we will need: We need to distinguish the deep-regret box from the ordinary (mild) regret box: My regret that I\* was a little late for the meeting is not in my deeply-regret box.

<sup>26</sup>Many of the ideas in this paper have roots in BAKER 2013.

## References

- BAKER, L. R. 1991, "Has Content Been Naturalized?", in *Meaning in Mind: Fodor and His Critics*, ed. by B. LOEWER and G. REY, Blackwell Publishing, Oxford, 17–32.
- 2007, *The Metaphysics of Everyday Life: An Essay in Practical Realism*, Cambridge University Press, Cambridge.
- 2011, "How to Have Self-Directed Attitudes", in *Self-Evaluation: Affective and Social Grounds of Intentionality*, ed. by A. KONZELMANN ZIV, K. LEHRER, and H. B. SCHMID, Springer, Dordrecht, 33–43.
- 2013, *Naturalism and the First-Person Perspective*, Oxford University Press, New York.
- CASTAÑEDA, H. N. 1966, "He: A Study in the Logic of Self-Consciousness", in *Ratio*, 8, pp. 130–157.
- 1967, "Indicators and Quasi-Indicators", in *American Philosophical Quarterly*, 4, pp. 85–100.
- CHALMERS, D. J. 1995, "Facing Up to the Problem of Consciousness", in *Journal of Consciousness Studies*, 2/3, 200–219.
- CHURCHLAND, P. 1988, *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, The MIT Press, Cambridge MA.
- DENNETT, D. C. 1995, *Darwin's Dangerous Idea*, Simon and Schuster, New York.
- ELDREDGE, N. 2000, *The Triumph of Evolution*, W.H. Freeman, New York.
- HUME, D. 1968, *A Treatise of Human Nature*, Clarendon Press, Oxford.
- KIM, J. 2000, *Mind in a Physical World*, The MIT Press, Cambridge MA.
- NAGEL, T. 1979, "Subjective and Objective", in *Mortal Questions*, Cambridge University Press, Cambridge, 196–213.
- 1986, *The View from Nowhere*, Oxford University Press, Oxford.
- PASNAU, R. 2002, *Thomas Aquinas on Human Nature*, Cambridge University Press, Cambridge.
- PERRY, J. 1979, "The Problem of the Essential Indexical", in *Noûs*, 13, pp. 3–21.
- SEARLE, J. R. 1992, *The Rediscovery of the Mind*, The MIT Press, Cambridge MA.
- 2000, "Consciousness", in *Annual Review of Neuroscience*, 23, 557–578.