



Anita Pacholik-Żuromska

## False Re-Representations in Self-Knowledge

**Abstract.** Self-knowledge (SK) is a natural ability of the human cognitive system and is defined as a complex re-representation of knowledge subject has about her own internal states. It is composed of two basic representations: the representation of I and the representation of the experienced state. SK has a propositional (i.e. language-like) form and can be expressed in the form of self-reports like “I believe that I believe that  $p$ ”. It has then the form of a second-order belief which, as a re-representation, generates the following problem concerning false representations: if the first-order representations of the I or of experienced states can be false like in cases of self-illusions or disorders like somatoparaphrenia, then how can error in self-cognition inherited by the re-representation (i.e. by SK) be avoided?

I claim that the re-representational model of SK is a useful but artificial tool for the description of the components of SK; however, it is misleading to think of it as capturing the truth about SK. I propose here a representational model of SK which reduces those beliefs constituting SK to first-order beliefs which are sufficient as constituents of SK.

**Keywords:** self-knowledge; representations; misidentification; propositional attitudes; self-model; applied logic

### Introduction

Self-knowledge (SK) is classically defined in philosophy as the subject’s knowledge about her conscious mental states (see, e.g., Descartes, 1637/2009; Boghossian, 2008; Peacocke, 1999). It has a propositional (i.e. language-like) form and can be expressed in the form of self-reports like “I believe that I believe that  $p$ ” (where “ $p$ ” stands for the content of the experienced state). The same definition is given by psychologists although the term “SK” has been replaced by the term “meta-consciousness” or “meta-awareness” (Schooler et al., 2011). According to the psy-

chological definition, meta-consciousness or meta-awareness is knowledge about our introspective experiences (Schooler and Schreiber, 2004, p. 17). Both psychology and philosophy in speaking either about “meta-consciousness” (Schooler and Schreiber, 2004) or about second-order belief (Peacocke, 1992, 1999) legitimate SK as being re-representational.<sup>1</sup>

SK as a complex re-representation is composed of two basic representations: the representation of the I and the representation of an experienced phenomenal state, such as pain or belief. On the one hand, SK has a special epistemological status because it rests on first-person authority legitimated by the direct access of the subject to her own mental states. Only in this way SK can be infallible, incorrigible and self-intimating (Guttenplan, 1994, p. 91). On the other hand, SK has to be a species of knowledge, then it must be a true and appropriately justified belief (Boghossian, 2008).<sup>2</sup> Cognition that leads to SK must be characterized by conditions of knowledge; i.e. the content of the knowledge-state must have satisfaction conditions. Content is here understood as the proposition that  $p$ . If  $p$  refers to the actual state of affairs i.e. such that corresponds with reality, then  $p$  is satisfied.

It seems that both conditions of SK are mutually exclusive. The first one excludes representationalism because it seems that between the subject and her state there should be no intermediaries, such as representations. SK as given directly should be presented, not represented to the subject. The second assumes representationalism because for SK to function as justified true belief, it must have a propositional form where the subject stands in the relation of belief (holding true) to a proposition  $p$  that is the content of the belief. The contents of mental states such as beliefs bearing information about the world must be somehow represented to the subject and hence the content of SK bearing information about subject’s mental state must be somehow re-represented to her.

---

<sup>1</sup> It is important to note that not every self-representational system is able to have SK, but every system able to have SK is self-representational.

<sup>2</sup> The precise analysis of knowledge is not of concern in this paper and so issues about how to understand the concept of justification, including the internalist/externalist distinction and whether justification is necessary for knowledge and whether, if so, justification can be understood in an externalist way, are of no concern. I shall understand “justification” to be that which is necessary to transform a true belief into knowledge. In so doing, I will not be engaging with so-called ‘knowledge-first’ accounts which reject the traditional view that the knowledge is a type of true belief (see, e.g., Williamson, 2000).

### The immunity of SK to error

The problem with SK is that although it should be immune to error being, through introspection, infallible and incorrigible (Guttenplan, 1994; Shoemaker, 1994), the empirical evidence shows that sometimes a self-representational system with SK — a subject of mental states — may be wrong about her experienced states. For example, consider the pathological case of a subject who has trouble judging whether the state of experience belongs to her, as in the case of somatoparaphrenia (Lane and Liang, 2009; Rosenthal, 2010). Experiments with self-illusions have confirmed the fragility of SK by showing that the first-person perspective can be manipulated (cf. Limanowski 2014; Ionta et al., 2011, p. 372; Lenggenhager et al., 2007; Blanke and Metzinger, 2009; Aspell et al., 2012). In all these cases, either the representation of the owner (the representation “I”) or the representation of the subjective experience, although vivid for the subject, are delusive- they are not objectively true. For example, the experience of mistaken self-location is exactly what impairs SK. The subject is of course sure that she experiences a state with the content  $p$ , but  $p$  is false and what’s more, it cannot be falsified from the first-person perspective, because sometimes the subject simply does not know that she is deluded. This was very clearly shown by the experiments inducing a “full-body illusion”, where a subject sees her avatar in front of her whilst being simultaneously stroked on her back before being moved a few meters back and asked to return to her previous place. In most cases the subjects come back to the place where they saw their avatar and not where they were really standing (cf. Lenggenhager et al., 2007). The content of experienced state was aptly summarized as: “I am there, where I feel myself” (Aspell et al., 2012). The problem was that the feeling was located in the body of the avatar rather than in their own.

Taking into account the above cases, the following questions arise:

- (1) Is SK really knowledge if the self-representational system can believe that it is in the particular state whilst it is *de facto* wrong about it?
- (2) Can the self-representational system building a computational self-model correct this error in the frame of a closed loop of self-reference?
- (3) If (2) receives a positive answer then the correction must happen at a higher level of representation, thus generating the next question: can re-representation correct the error raised at a lower representational level?

In my opinion, for SK to be reliable, it is required that it acts as an indicator of false states formed at the level of phenomenal self-experience: i.e. that the subject can be in a position to know that she is experiencing a false state (such as the full-body illusion). In order to do this, a subject has to be equipped with some cognitive tools enabling the recognition of such errors as misidentification of the subject's own physical and mental states or misidentification of ownership. Such tools cannot be constituted entirely internally as the error can arise at the level of self-specifying processes and be repeated at subsequent levels of information-processing until it appears in the form of a conscious experience. In that case a subject cannot know just from introspection that her experience is false.

### **Why is representationalism bound to the propositionality of SK?**

In the traditional representational account, SK should be immune to error through misidentification by virtue of the possession conditions of the concepts involved in forming first-person judgments (Peacocke, 1992, 1998; Bermúdez 2012). The representational character of SK is established precisely by its propositional structure. Propositional attitudes are expressed in sentences containing the I-clause and the that-clause forming a relation between a subject, an attitude and a judgment in a logical sense. Every propositional attitude has an intentional character, expressing the subject's attitude towards a given judgement. Philosophers supporting representationalism such as Fodor (1998) would say that a propositional attitude is a relation between a subject and a representation, where the minimal constituents of representation are modes of presentation. This last term Fodor borrows from Frege (*Art des Gegeben-seins*). According to Fodor, a thought is a mental representation which carries intentional content that it inherits from its constituents, which are concepts. Mental representations are part of the judgments that are the objects of propositional attitudes. Hence, beliefs, i.e. propositional attitudes are layered; see Figure 1.

On the other hand, Peacocke (1992) claims that the constituents of propositional content are concepts that are individuated through their possession conditions. These possession conditions together with a determination theory show how a given concept's semantic value is fixed and guarantee the rational sensitivity of a subject. The determination theory allows a subject to form true beliefs and to make truth-preserving infer-

## First-order belief:

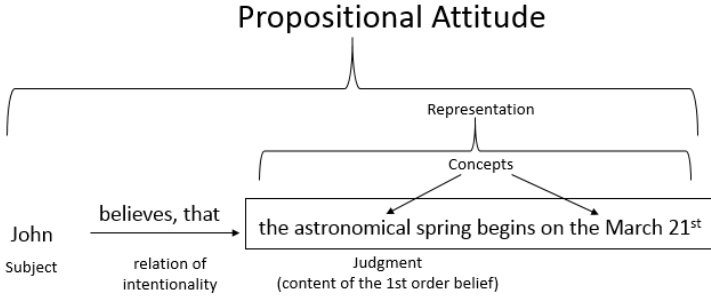
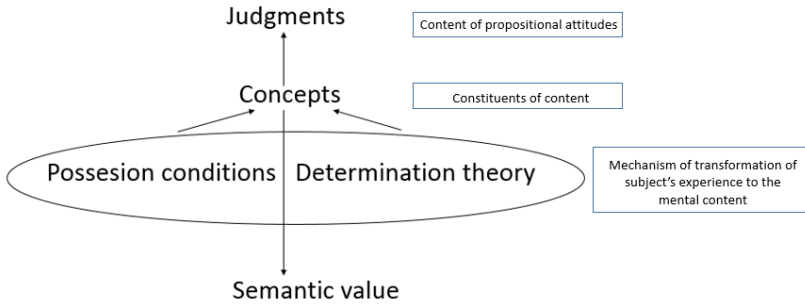


Figure 1.

ences in the belief-forming practices included in the concept's possession conditions determining the correct grasp of the concept and allowing the subject to give correct judgments containing this concept (cf. Peacocke, 1992). Possession conditions and determination theory are just general notions of the mechanism of the transformation of a subject's experience into the mental content expressed in a belief. Fodor, as an internalist, would not speak about possession conditions because they underline the necessity of subjects' interaction with the outer world to form the content of beliefs. Instead he uses the notion of the causal power of mental representations by virtue of their intrinsic properties (Fodor, 1985). But what is in this case extremely interesting is that both internalists and externalists employ a Fregean semantics (cf. Frege, 1892/1994). According to Frege, names (*Eigennamen*) have a sense (*Sinn*) and a reference (*Bedeutung*). The sense of a name is a concept whose content is the way an object is given to us, or its mode of presentation (*Art des Gegeben-seins*). The reference of a name is a real object. The discussion between internalists and externalists concerns the question whether the senses (intensions) are in the head and can determine the extension of a name or are they the way in which an object is given to a subject (metaphorically speaking, they are "stuck" to the external things) and they determine the mental content from the outside, as Peacocke would say (1992).

The whole process of the formation of a propositional attitude lies in the fact that to concepts are assigned logical functions that transform objects into logical values, so that the subject having a given concept will make true judgments. Thus, the subject has a true belief, if the

semantic value of the concept — the minimal constituents of this belief — is a function from objects to truth-values (cf. Peacocke, 1992, p. 21).



For somebody who attributes propositional attitudes with a particular content to herself or others a kind of so-called “rational sensitivity” is required. Fodor (1985, p. 77) formulates it in these words:

It is, in short, untendentious that people regularly account for their voluntary behaviour by citing beliefs and desires that they entertain; and that, if their behaviour is challenged, they regularly defend it by maintaining the rationality of the beliefs (‘Because it says it’s going to Clapham’) and the probity of the desires (‘Because it’s nice visiting Auntie’).

The mechanisms of the transformation of the folk psychological attitudes into coherent mental content rely on the relations between representations:

For any organism *O* and for any proposition *P*, there is a relation *R* and a mental representation *MP* such that: *MP* means that (expresses the proposition that) *P*; and *O* believes that *P* iff *O* bears *R* to *MP*. (And similarly, *R* desires that *P* iff *O* bears some different relation, *R'*, to *MP* and so forth). (Fodor, 1985, p. 88)

Peacocke, on the other hand, describes the “rational sensitivity” as sensitivity to the satisfaction of the possession conditions for the concepts in the content attributed:

For any mode of presentation *m*, we can say that a use of *m*, in a particular belief ‘*m* is *F*’, is representationally dependent if there is a content ‘*m* is *F*’ having the properties which result from clauses (i)–(ii) above by substituting ‘*m*’ for ‘*I*’. That is, the content ‘*m* is *F*’ must be the content of one of the thinker’s current representational mental states;

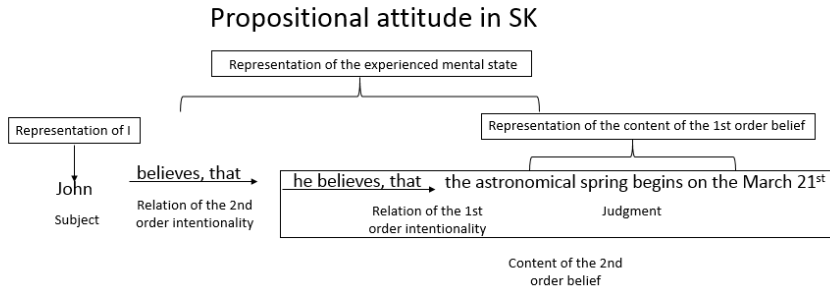
and the thinker must form the belief ‘m is F’ because he is in the mode of taking such states at face value. A use of a perceptual-demonstrative such as ‘that car’, in an ordinary perceptual demonstrative belief like ‘That car is travelling fast’, is representationally dependent.

(Peacocke, 1999, p. 266)

The above-mentioned mechanisms work for the formation of the mental content of first-order beliefs, but they can be extrapolated to the mechanisms of SK.

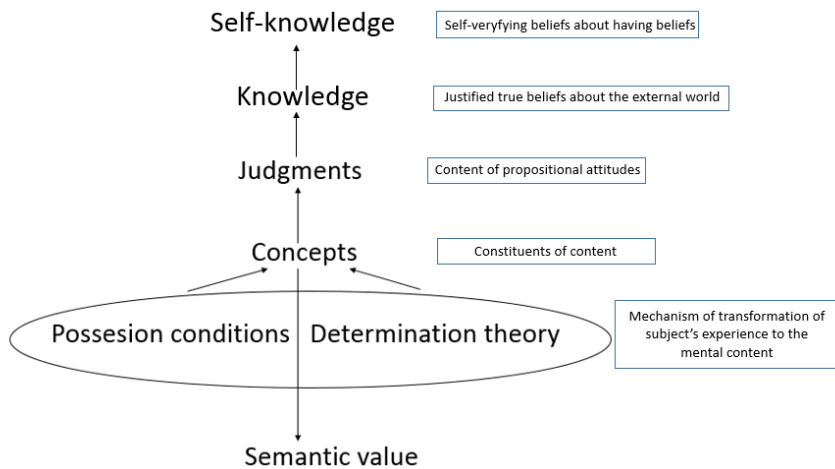
In SK to ascribe an attitude to herself a subject needs to have two abilities: to have a concept of the self – the “I” – to ascribe the attitude to herself as herself, and to have a concept of the attitude like a belief or desire. In other words, to form an SK-judgment a subject has to have the ability to form second-order beliefs and hence she needs to have access to concepts involved in the content of the beliefs of this kind. The propositional attitude in SK expressed by second-order beliefs about the mental states of the subject is *de facto* an attitude about an attitude.

### Second-order belief:



SK arises through transitions made by the subject between the content of the first-order beliefs to the content of the second-level belief using the proper inferences. It is allowed by the very nature of such a content because the content of SK holds i.e. inherits the truth-value from the primary content of the first-order belief (cf. Peacocke, 1992). The mechanism of this inheritance rests on the inferential role semantics expressed in the Generality Constraint formulated by Peacocke after Gareth Evans, which explains the productivity of beliefs (Peacocke, 1992, pp. 42, 46).

When a subject thinks “I am F”, where the I is the object, then we are dealing with something with the classical form of a propositional attitude “I believe that *p*”. It means that the propositional form of



attitudes involves the I as object and that what comes after “that” is a predicate. However, such an approach to the propositionality of SK is unacceptable because it generates mistakes in self-ascriptions. It is possible that someone is F but she misidentifies herself as the person who is F (Bermúdez, 2012, p. 124.). In other words, the connection between the subject and object in SK is not established. Although the perceptual model of SK allows us to explain problems with first-person authority, by the fact that either the subject’s access to the object (i.e. internal states of the subject) is disturbed or these states are not clear, and therefore the object itself is inaccessible, this model does not help to solve the problem of SK as knowledge. In other words, it provides an explanation of an abnormal situation, where a subject misidentifies her own states, but it does not show a way out of error in self-cognition, for example by giving the subject the possibility of self-correction or even of judging that she can be mistaken about her own mental states. Some attempts to avoid the perceptual model of SK are introduced below.

### Examples of the contemporary models of SK

Recent approaches to the issue of SK are strongly linked to the conception of an active inference. Positions on active inference can be externalist by saying that the interaction with the world (action performance) serves as a tester of sensory input (Seth, 2015) or internalist by saying



that active inference does not need a non-neural body for predictive control (Metzinger, 2017). Predictions made by the brain allow the subject to anticipate the current states of the world on the basis of sensory input for the purpose of performing an appropriate action (Friston, 2009). If the external sensory input fits the hypothesis (prediction) it is improved, whereas when it is different the hypothesis is modified. These hypotheses can be understood as representations as *de facto* they represent the expected states of affairs (cf. Hohwy, 2012). The predictions are therefore a model of a probable (expected) situation (cf. Limanowski, 2017).

### A representational model of SK<sup>3</sup>

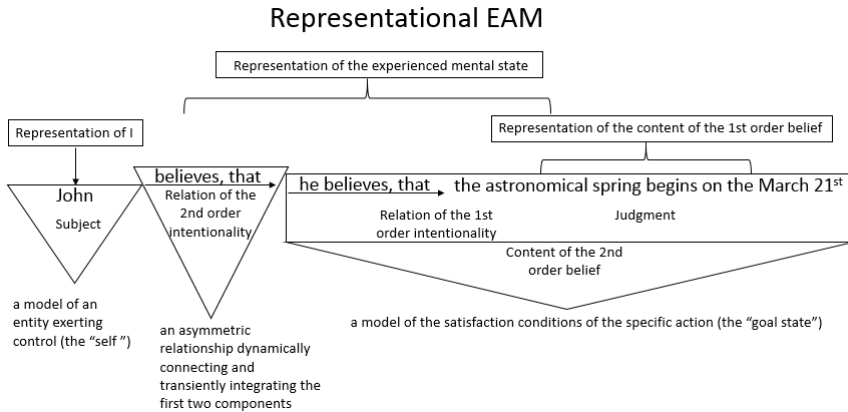
According to Metzinger (2017) active inference can not only be used to predict the states of the world but also to explain the process of self-modelling. In other words, the brain represents (puts forward hypotheses) not only (about) the external state of affairs but also (about) its owner. Such predictions are made in the epistemic agent model (EAM) formed at the level of conscious processing and representing its owner as an individual capable of autonomous epistemic self-control, i.e. monitoring and voluntary modification of own mental states (Metzinger, 2017, p. 8). This model is based on the concept of minimal phenomenal selfhood – the subjective experience of being a self, independent of explicit cognition. In other words, minimal phenomenal selfhood is a fundamental element of EAM because to ascribe mental states to oneself a subject needs to refer to herself with “I”.

The EAM is an effect of the relation between the self and the intentional goal, with the goal being understood as an inner mental object (Metzinger, 2017). It is independent of any external factor, represents the ability to become an active cognizer and generates the subjective experience of ownership of mental states (Metzinger, 2017). The active cognizer is defined as an epistemic agent, whilst mental agency is under-

---

<sup>3</sup> The precise analysis of knowledge is not of concern in this paper and so issues about how to understand the concept of justification, including the internalist/externalist distinction and whether justification is necessary for knowledge and whether, if so, justification can be understood in an externalist way, are of no concern. I shall understand “justification” to be that which is necessary to transform a true belief into knowledge. In so doing, I will not be engaging with so-called ‘knowledge-first’ accounts which reject the traditional view that the knowledge is a type of true belief (see, e.g., Williamson, 2000).

stood as predictive control of effective connectivity — the epistemic value of internal states contained in self-model (Metzinger, 2017).



EAM does not need any external (non-neural) body to be present (Metzinger, 2017). On the one hand, EAM allows for self-consciousness, whilst on the other it is the model of knowing the self (Metzinger, 2017, p. 8). What constitutes EAM as a model of SK are voluntary mental actions such as attentional agency, whereas the epistemic value in EAM concerns the future states of an organism's self-model (Metzinger, 2017, p. 15). Satisfaction conditions for a predicted internal state would be specified in this very prediction. The hypothesis that EAM can be entirely neural, i.e. it does not involve the participation of any external elements which would be necessary for self-modelling, leads to internalism and the Representational Theory of Mind, according to which the representational content of mental states is determined only by internal states of an organism i.e., by the other representations (Fodor, 1987, 1991). Thus even a hypothetical brain in a vat would be able to create an EAM. Mental actions result from a wholly internal, sub-personal (i.e. neural) information-processing and do not extend beyond the body of an organism (Metzinger, 2017, p. 2). Hence it is plausible that a hypothetical brain in a vat, deprived of any contact with the external world, being only subject to internal neural information-processing would be able to create its own model as the subject of the states, which she will ascribe to herself.

But could brains in a vat truly say that they are brains in a vat? According to EAM they could represent themselves as brains in a vat but

also as something else as long as they actively make inferences solely on the basis of neural stimulation. What they cannot do is to test whether their representations correspond to reality. However, even the statement “I am a brain in a vat” would not be true because of the lack of external reference — external satisfaction conditions. Mental states (intensions) do not determine extensions (Putnam, 1981, pp. 1–21). The well-known thought experiment of Putnam refutes the necessary connection between representations and their objects.

### **An anti-representational model of SK<sup>4</sup>**

In order to avoid the requirement of building models based on the notion of representation, some researchers such as Anil Seth have turned to cybernetics, replacing the role of representation by the loop of the brain-body-world interaction (Seth, 2015, p. 15). Seth rejects the traditional account of the role of representations as intermediaries between internal states and the world. His view refers to the conception of so-called *en*-cognition, which should be embodied, extended, enactive and external conditioned. Seth says that “It is enactivism that is most explicitly anti-representationalist. While enactive theorists might agree that adaptive behavior requires organisms and control structures that are systematically sensitive to statistical structures in their environment, most will deny that this sensitivity implies the existence and deployment of any “inner description” or model of these probabilistic patterns” (Seth, 2011, p. 16).

Seth argues that “perception emerges as a consequence of a more fundamental imperative towards homeostasis and control, and not as a process designed to furnish a detailed inner “world model” suitable for cognition and action planning” (Seth, 2011, p. 1385). This is of course a view which escapes from the Cartesian trap of internalism because the necessary element of the model of SK is here an external element, i.e. subject’s environment. Hence, the role of the connection of an organism with the world as being necessary for cognition is somehow obvious but not trivial.

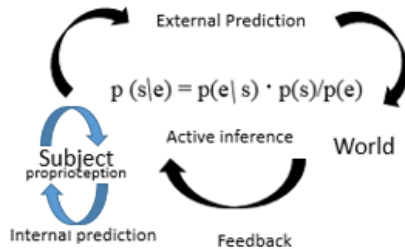
The combination of PP which “takes into account external and internal perception, instantiates a process of interoceptive inference” (Seth, 2015).

---

<sup>4</sup> An extended description and analysis of this model in the light of externalism is given in (Pacholik-Žuromska, 2019).

This concept provides a natural way of thinking of the neural substrates of emotional and mood experiences, and also describes a common mechanism by which interoceptive and exteroceptive signals can be integrated to provide a unified experience of body ownership and conscious selfhood. (Seth, 2015, p. 3)

A model of SK involving anti-representationalism will not by definition make room for SK ordinarily understood as knowledge of a proposition; rather, it will turn out to be a kind of a model of bodily self-consciousness. It will also dispense with the concept of intentionality in the traditional account as aboutness and introduce instead the idea of a goal direction.



Unfortunately, by building an anti-representational model of SK, we throw the baby out with the bathwater. How can true justified beliefs be formed if there is no place for propositionality, which is necessarily bound to representations?

### **A proposed solution on the basis of the distinction between false representation and misrepresentation**

A way of to solve the problem of the propositionality of SK and preserve immunity to error is to put forward a kind of connection between representationalism and anti-representationalism. First of all, what is to be avoided in representationalism is the perceptual model of SK. Some suggestions have been given by Sydney Shoemaker, who says that we do not identify ourselves via experiences of mental states, but we identify our states via ourselves, i.e. via whom they belong to (Shoemaker,

1994).<sup>5</sup> An identification requires a prior reference to oneself, and this means that self-identification is something primary to the identification of one's own state. Therefore, the knowledge expressed by "I feel pain" does not concern the content of a given state, but the fact that it is "Me and no one else who is in pain" (Lane, 2011).

The attitude "I believe that  $p$ " can be transformed into "I am F" where "I" is the subject and "believe that  $p$ " is a predicate F. If "I believe that  $p$ " can be transformed into "I am F", where for "F" stands the belief that  $p$ , then SK, as a second-order belief, can be reduced to "I am F" according to the following reasoning:

- (i) The second-order belief is composed of the judgement about one's own attitude: "I believe that I believe that  $p$ " ,
- (ii) "I believe that  $p$ " has the form "I am F",
- (iii) From (i) and (ii) follows: "I believe that I am F",
- (iv) From (ii) and (iii) follows: "I am F" because the second-order belief can be also expressed in the form of a predicate.

Ergo: "I am F" already expresses knowledge about a subject's mental state and we do not need the second-order belief.

The re-representation in SK should be considered as a special kind of representation expressed in propositional form depicted in the above reasoning. It is not to thought of as a relation between the subject and a propositional content (the I and the judgement  $p$ ). It is rather a special kind of unified predication. Following Shoemaker, it is the predication of a certain kind of complex experience to an already-identified self. At least in such an account we avoid error in self-identification: I can be wrong about the content of an experienced state, but I cannot be wrong that it is me who experiences it.

The problem arises when we distinguish between the false representation of and the misidentification of mental states.<sup>6</sup> The main mistake about the propositional (i.e. representational) character of SK is made

---

<sup>5</sup> Shoemaker's idea was based on an idea from William James who distinguished between two types of self: present (I) and narrative (Me). The first of them is the experiencing self—the subjective one; the second is me as the object to which mental and physical predicates are assigned (James, 1890/1950). The self in SK is the subjective, experiencing one, not the narrative me to whom we relate through specific descriptions. The first one accounts for the self in SK and is a guarantee of resistance to error in self-cognition.

<sup>6</sup> Arguments for such a distinction are given in Dretske (1981).

by the distinction into two kinds of first-person-reports: phenomenal reports linked with misidentification and propositional reports linked with representation.

- (1) I feel pain,
- (2) I believe that astronomical spring begins on March 21<sup>st</sup>.

(1) is a report of phenomenal experience. (2) is a belief (i.e. propositional attitude) expressing an attitude of the subject toward a judgment. Involved in the structure of SK they will have the form:

- (1') I believe that I feel pain,
- (2') I believe that I believe that astronomical spring begins on March 21<sup>st</sup>.

I claim that the distinction between reports (1') and (2') is illusory. They are the same kind of report, namely reports about a subject's experiences which can be expressed in the form "I am F" and hence they fall under the reasoning i-iv. Consequently, I claim that beliefs reported in SK as its content are a kind of experience despite their different structure from sensory perceptions. They are still inner states of a subject — of various types — but in SK it makes no difference. A subject having a first-order belief experiences them both as belonging to her (of course in so far as she is conscious of them). SK is knowledge about the conscious inner states of a subject regardless of their contents, and so the distinction between (1') and (2') is invalid only in the traditional approach to SK. Both the first-order beliefs and the feeling of pain make the intentional content of the beliefs of the second order, i.e. SK.

It should be noted here that the conception of beliefs as experiences is different from the view that SK reports have the same status as avowals (cf. Bar-On, 2004). Avowals are also expressions providing information about the subject's inner state; however, they do not make explicit the content of SK but rather only stand for the content. "Ouch!" can be an expression of pain but SK in form of a belief contains also an attitude toward its content. SK-states are experiences but not avowals because the content of the belief \*is\* made explicit. When I believe that I am in pain, that belief is part of the identity of the experience.<sup>7</sup> Hence, SK cannot be reduced to the experience making up its content because SK is about the experience.

---

<sup>7</sup> I thank Matthew Carmody for helping me clarify my position in this matter.

A possible objection against the proposed account is the concern that the form “I am F” expressing SK does not mean that the content of SK is only *presented* in self-consciousness (and so how anti-representationalism wants it to be) and not *represented*, as it should be in the proper account of SK involving second-order beliefs (beliefs about beliefs). This objection can be refuted by saying that expressions such as first-order beliefs that inform about other experiences already have a propositional form. The first-person statements “I feel pain” or “I believe that it is raining” contain at the same time information about the content of the experienced state and the information about the content of SK. In other words, they have an implicit structure of re-representation with reference to the outer object like rain or the cause of pain (for example in a limb, which can be false like in case of phantom pain) but at the same time refer to the subject as having this belief (which is cannot be false, because the subject really feels pain although the cause of the pain is misidentified).

Another objection could be that this account of SK is problematic because it reminds us of the deflationary theory of truth where the word “true” is a tool of assertion. By saying that *p is true* we express the same judgment as just by saying that *p* (cf. Wright, 1999). Hence the second-order beliefs in SK could be understood as such tools of assertion, which could be omitted, because the first-order beliefs already express SK. However, the second-order beliefs are here rejected for reasons other than those given in the deflationary theory of truth: the first-person clauses are asserted by virtue of the privileged access of the subject to her own mental states and already contain all the information needed for SK. There is no need to multiply this information in the second-order belief.

A third objection could be that this approach to SK entails that every creature experiencing some states like pain or fear already has SK. The answer is that a subject not only has to have an ability to experience a state but also she has to have the ability to represent herself as a subject of this experience and to represent the experienced state as belonging to herself.

The fourth objection is the most serious: does the proposed account of SK offer a way out from the Cartesian trap of an internal loop of the subject’s false beliefs about her own mental states? Here the answer is positive. This model of SK, which involves only first-order beliefs has an external character which means that it contains essentially an external

element from subject's environment, i.e. the confrontation with beliefs of others. This is what is borrowed from anti-representationalism. The rational sensitivity of the subject is constituted not only by the internal mechanisms of SK but also by interaction with the social and physical environment. A subject must be open to the possibility that she can be wrong about her own mental states. The ability to cast doubt on her judgments about her own mental states is an element of rational sensitivity. Information from others such as "You are experiencing a pain in your leg but it is a phantom pain because your leg has been amputated" are integrated in the belief-system of the subject, become a part of SK and change the content of SK. Such external information sometimes changes also the experience of the subject such as in the case of the mirror box (cf. Ramachandran, 2009) and sometimes it does not, but it changes the attitude toward the experience. A subject knows then that her experience is false and modifies her first-order beliefs by not saying anymore "I feel pain", but by saying "I feel phantom pain".

### Summary

The re-representational form of SK is depicted by a two-level model of propositional attitudes. This model, however, is misleading because it presents SK either in a perceptual model or as being about beliefs of the same sorts as those about the world external to the subject. It is an artificial tool for the description of SK which has little reference to reality. The preferred model, developed in this paper as an alternative to the traditional model of SK, because it shows more precisely just what SK is. A subject consciously informing us that she believes, experiences or desires *p*, already knows that she does. It does not matter whether *p* is objectively true. Sometimes, such as in the aforementioned cases, a subject can be deluded. For her the experienced state is true. But the subject also has the ability to be rationally sensitive to her beliefs about experienced mental states, which allows her to internalize and integrate the beliefs of others, such as those of experts, or beliefs formed on the basis of the perception of external objects, and in this way to modify the content of SK.



**Acknowledgments.** I would like to thank the anonymous reviewers for their helpful and valuable comments and efforts towards improving my manuscript. I also thank Matthew Carmody for the professional language correction and the suggestions helping me to express my thoughts more precisely. His help is a good example of how others can sometimes know better what we ourselves think.

## References

- Aspell, J. E., B. Lenggenhager and O. Blanke, 2012, “Multisensory perception and bodily self-consciousness: From out-of-body to inside-body experience”, Chapter 24 in M. M. Murray, M. T. Wallace (eds.), *The Neural Bases of Multisensory Processes*, Boca Raton (FL): CRC Press, Taylor & Francis. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK92870/>
- Bar-On, D., 2004, *Speaking My Mind. Expression and Self-Knowledge*, Oxford: Clarendon Press.
- Bermúdez, J. L., 2012, “Memory judgments and immunity to error through misidentification”, *Grazer Philosophische Studien* 84: 123–142.
- Blanke, O., and T. Metzinger, 2009, “Full-body illusions and minimal phenomenal selfhood”, *Trends in cognitive sciences* 13: 8–13. DOI: [10.1016/j.tics.2008.10.003](https://doi.org/10.1016/j.tics.2008.10.003)
- Boghossian, P. A., 2008, *Content and Justification. Philosophical Papers*, Oxford: Oxford Clarendon Press.
- Case, L., R. Abrams and V. S. Ramachandran, 2010, “Immediate interpersonal and intermanual referral of sensations following anesthetic block of one arm”, *Archives of Neurology* 67(12): 1521–1523.
- Descartes, R., 1637/2009, *Discourse on the Method of Rightly Conducting the Reason and Seeking Truth in the Field of Science*, SMK Books.
- Dretske, F., 1981, *Knowledge and the Flow of Information*, Cambridge, Massachusetts: The MIT Press.
- Fodor, J. A., 1985, “Fodor’s guide to mental representation: The intelligent Auntie’s vade-mecum”, *Mind* 94 (373): 76–100.
- Fodor, J. A., 1987, *Psychosemantics*, Cambridge, MA: The MIT Press.
- Fodor, J. A., 1991, “A modal argument for narrow content”, *Journal of Philosophy* 88: 5–26.
- Frege, G., 1892/1994, “Über Sinn und Bedeutung”, in G. Patzig (ed.), *Funktion, Begriff, Bedeutung*, Göttingen, Vandenhoeck & Ruprecht.

- Friston, K. J., J. Daunizeau and S. J. Kiebel, 2009, “Reinforcement learning or active inference?”, *PLoS ONE* 4, e6421. DOI: [10.1371/journal.pone.0006421](https://doi.org/10.1371/journal.pone.0006421)
- Guttenplan, S., 1994, “First person authority”, in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*, Oxford: Blackwell.
- Hohwy, J., 2012, “Attention and conscious perception in the hypothesis testing brain”, *Frontiers in Psychology* 3. DOI: [10.3389/fpsyg.2012.00096](https://doi.org/10.3389/fpsyg.2012.00096)
- Ionta, S., L. Heydrich and B. Lenggenhager, 2011, “Multisensory mechanisms in temporo-parietal cortex support self-location and first person perspective”, *Neuron* 28: 263–374.
- Lane, T., and C. Liang, 2009, “Higher-order thought and pathological self: The case of somatoparaphrenia”, *Analysis* 69 (4): 661–668. DOI: [10.1093/analys/anp094](https://doi.org/10.1093/analys/anp094)
- Lenggenhager, B., T. Tadj, T. Metzinger and O. Blanke, 2007, “Video ergo sum: Manipulating bodily self-consciousness”, *Science* 317: 1096–1099.
- Limanowski, J., 2014, “What can body ownership illusions tell us about minimal phenomenal selfhood?”, *Frontiers in Human Neuroscience* 8: 946. DOI: [10.3389/fnhum.2014.00946](https://doi.org/10.3389/fnhum.2014.00946)
- Limanowski, J., 2017, “(Dis-)attending to the body – action and self-experience in the active inference framework”, in T. Metzinger and W. Wiese (eds.), *Philosophy and Predictive Processing*, 18, Frankfurt am Main: MIND Group. DOI: [10.15502/9783958573192](https://doi.org/10.15502/9783958573192)
- Metzinger, T., 2017, “The problem of mental action: Predictive control without sensory sheets”, in T. Metzinger and W. Wiese (eds.), *Philosophy and Predictive Processing*, 19, Frankfurt am Main: MIND Group. DOI: [10.15502/9783958573208](https://doi.org/10.15502/9783958573208)
- Pacholik-Żuromska, A., 2019, “Modele jaźni w samowiedzy”, *Filozofia Nauki* 27 (3) [107]: 55–76. DOI: [10.14394/filnau.2019.0017](https://doi.org/10.14394/filnau.2019.0017)
- Peacocke, Ch., 1999, *Being Known*, Oxford: Clarendon Press.
- Putnam, H., 1981, “Brains in a vat”, pages 1–21 in *Reason, Truth and History*, Cambridge: Cambridge University Press.
- Ramachandran, V. S., B. Krause and L.K. Case, 2011, “The phantom head”, *Perception* 40 (3): 367–370.
- Ramachandran, V. S., and E.L. Altschuler, 2009, “The use of visual feedback, in particular mirror visual feedback, in restoring brain function”, *Brain* 132 (7): 1693–1710.
- Rosenthal, D. M., 2010, “Consciousness, the self and bodily location”, *Analysis* 70 (2): 270–276. DOI: [10.1093/analys/anp167](https://doi.org/10.1093/analys/anp167)

- Shoemaker, S., 1994, “Self-knowledge and “inner sense””, *Philosophy and Phenomenological Research* LIV (2): 249–314.
- Schooler, J. W., and C. A. Schreiber, 2004, “Experience, meta-consciousness and the paradox of introspection”, *Journal of Consciousness Studies* 11 (7–8): 17–39.
- Seth, A. K., 2015, “The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies”, in T. Metzinger and J. M. Windt (eds.), *Open MIND*, Cambridge, MA: The MIT Press.
- Williamson, T., 2000, *Knowledge and its Limits*, Oxford: Oxford University Press. DOI: [10.1093/019925656X.001.0001](https://doi.org/10.1093/019925656X.001.0001)
- Wright, C., 1998, “Truth: A traditional debate reviewed”, *Canadian Journal of Philosophy* 28 (sup1): 31–74.

ANITA PACHOLIK-ŻUROMSKA  
Department of Cognitive Science  
Institute of Information and Communication Research  
Nicolaus Copernicus University in Toruń  
[apz@umk.pl](mailto:apz@umk.pl)