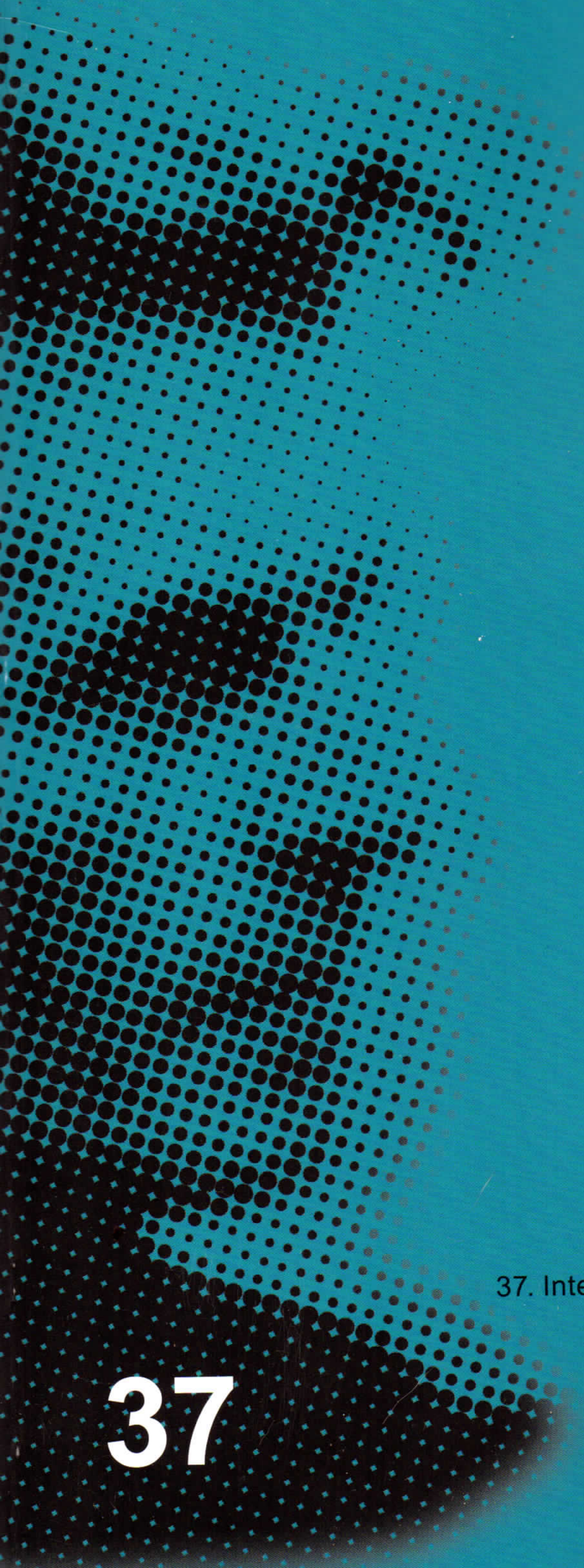

# Analytische und Kontinentale Philosophie: Perspektiven und Methoden

# Analytical and Continental Philosophy: Methods and Perspectives

Sonja Rinofner-Kreidl
Harald A. Wiltsche
Hrsg.

Beiträge
Papers

37. Internationales Wittgenstein Symposium
37th International Wittgenstein Symposium
Kirchberg am Wechsel
10. – 16. August 2014

37

# Analytische und Kontinentale Philosophie: Perspektiven und Methoden

# Analytical and Continental Philosophy: Methods and Perspectives

Beiträge der Österreichischen Ludwig Wittgenstein Gesellschaft
Contributions of the Austrian Ludwig Wittgenstein Society

Band XXII
Volume XXII

**Distributors**

Die Österreichische Ludwig Wittgenstein Gesellschaft
The Austrian Ludwig Wittgenstein Society

Markt 63, A-2880 Kirchberg am Wechsel
Österreich / Austria

*www.alws.at*

ISSN 1022-3398

Pre-Proceedings, Abstracta-Heft und Programm wurden mit Hilfe eines von Joseph Wang, Universität Innsbruck erarbeiteten Datenbankprogramms erstellt.
Kontakt: <joseph.wang@uibk.ac.at>

Pre-Proceedings, Book of Abstracts and Programme were produced using a database application developed by Joseph Wang, University of Innsbruck, Austria.
Contact: <joseph.wang@uibk.ac.at>

Visuelle Gestaltung: Sascha Windholz
Druck: Eigner Druck, A-3040 Neulengbach

# Transparency or Opacity of Mind?

Martin F. Fricke

Mérida, Mexico | martin_fricke@yahoo.co.uk

## Abstract

Self-knowledge presents a challenge for naturalistic theories of mind. Peter Carruthers's (2011) approach to this challenge is Rylean: He argues that we know our own propositional attitudes because we (unconsciously) interpret ourselves, just as we have to interpret others in order to know theirs'. An alternative approach, opposed by Carruthers, is to argue that we do have a special access to our own beliefs, but that this is a natural consequence of our reasoning capacity. This is the approach of transparency theories of self-knowledge, neatly encapsulated in Byrne's epistemic rule (BEL): If *p*, believe that you believe that *p* (Byrne 2005). In this paper, I examine an objection to Carruthers's theory in order to see whether it opens up space for a transparency theory of self-knowledge: Is it not the case that in order to interpret someone I have to have some direct access to what I believe (cf. Friedman and Petrashek 2009)?

Self-knowledge presents a challenge for naturalistic theories of mind. This is because self-knowledge seems to be especially secure, while not obviously sharing the features that confer security in knowledge of other things. Descartes thought that (some kind of) self-knowledge was the most certain knowledge to be had and therefore tried to ground all other knowledge on this first certainty. Today we have somewhat lost faith in such a foundationalist programme. Naturalism in epistemology could be characterised as the view that scientific knowledge, especially knowledge of the natural sciences, is more certain, at least in its totality, than any kind of foundation that philosophers might propose for it. It does not need such a foundation. The best that philosophy can do is to try and integrate whatever theories it cherishes into to the system of scientific knowledge.

It might be thought that this is exactly what contemporary analytic philosophy is trying to do when discussing self-knowledge. Whether cherished or not, self-knowledge concerning beliefs, desires and phenomenal states seems to be exceptionally secure; perhaps not completely infallible, as Descartes might have thought, but much more secure than any ordinary or, indeed scientific, knowledge of the world. At the same time, it is clear that it lacks any of the features that characterise scientific knowledge. It does not seem to be based on observation, inference, experiments, a large body of theoretical knowledge, confirmation by peers and so on. Some think that it is not based on anything and it certainly seems to be much more direct than any scientific knowledge. If that is so, then, in the context of naturalism, the phenomenon of self-knowledge is in need of explanation. We might not want to accord it any great importance for the rest of our theories, as did Descartes. But there is no doubt that its apparent possibility does need to be accounted for.

In this paper I shall look at two theories of self-knowledge that approach this problem in different ways. The first one, proposed by Peter Carruthers, stands in the tradition of Gilbert Ryle (1949). It says that our knowledge of our own propositional attitudes is acquired in the same way as our knowledge of the propositional attitudes of other people. We have a mindreading module that we can either apply to others or to ourselves and when we apply it to ourselves we acquire self-knowledge of our own propositional attitudes. So self-knowledge is not really that special and different; it is just nourished by a special wealth of data, since we are with ourselves all day long, gathering evidence for possible self-attributions, while having to work with more limited data when it comes to others.

The second theory is that suggested by Alex Byrne, which in turn is inspired by Gareth Evans's famous remark that I "answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*" (Evans 1982: 225). Theories that take their cue from this observation have been labelled transparency theories of self-knowledge because they regard the question of whether I believe that *p* as "transparent to" the question of whether *p*. Byrne tries to encapsulate this transparency in the following epistemic rule:

> (BEL) If *p*, believe that you believe that *p* (Byrne 2005: 95)

Following this rule, even merely *trying* to follow it, but getting one's facts wrong (i.e. it is not true that *p*), will produce true self-ascriptions of belief. It does not rely on perception, but rather on very simple inferential skills: to go from "*p*" to "I believe that *p*". Byrne thinks that this explains the special security of self-knowledge. There is so little that can go wrong. In addition to this "privileged access", (BEL) is also supposed to explain our "peculiar access" to our own beliefs. If I were to go from "*p*" to "*He* believes that *p*", I would be much more likely to make a false belief-ascription. So the kind of access described in (BEL) is peculiar to one's own mind. There is no equivalent kind of access to other people's minds. So in contrast to Carruthers's Rylean theory of self-knowledge, Byrne attempts to show that we do have a special kind of access to our own minds, differentiated not only by the amount of data from which we reason about ourselves, but also by the *method* through which we know ourselves, a method which is only applicable to ourselves. We might say that Byrne answers the challenge to naturalism not by denying the specialness of self-knowledge, but by showing that this specialness is a consequence of the normal powers of reasoning combined with a simple epistemic rule.

My own sympathies lie with Byrne's account. So for the rest of this paper I shall discuss an objection that followers of Byrne might put forward against Carruthers's theory.

The originality of Carruthers's approach does not, of course, lie in his Ryleanism, but in the way he defends it with the help of contemporary cognitive science. Central to this defence is a modular theory of mind with a "global broadcast architecture". The idea is that the mind consists of different specialised systems organised around a common workspace. The systems cannot communicate directly with each other, but only via messages that are globally broadcast in the workspace, thereby becoming access conscious. This setup resembles somewhat that of a bunch of specialists in different areas of science who are gathered around a blackboard and can only communicate by writing on the board. The crucial feature, for our purposes, of the common workspace is that it can only broadcast *sensory* messages, for example perceptual states, images or instances of inner speech. Decisions, judgments, beliefs, intentions or other propositional attitudes cannot be globally broadcast as such. They have to be expressed in sensory states such as images or speech first. Now, one of the systems constituting the mind is a mindreading faculty, used to attribute mental states to other people (or, presumably, nonhuman animals). According to Carruthers, it is this mindreading faculty which also provides us with knowledge of our own mental states. For this purpose, it has to use the information it receives via the common workspace of the mind. It does not have a general direct access to the subject's propositional attitudes because these are not globally broadcast, except when transformed into sensory data. And even if they are presented sensorily, the sensory data – what we hear someone, in this case ourselves, saying, for example – still have to be *interpreted* by the mindreading faculty to determine what propositional attitude is expressed by them.

If all of this is true, then the mindreading faculty cannot apply a procedure such as Byrne's epistemic rule (BEL). To apply the rule "If *p*, believe that you believe that *p*" we first have to know (or at least think that we know) that *p*. In other words, we have to have access to the content of the first-order belief to be attributed, i.e. we have to have access to what is believed. But if Carruthers is right, the mindreading system does not have a general access to what is believed. It only has access to what is perceived or imagined in some sort of way and thus globally broadcast in the workspace. (So if there were a procedure analogous to that encapsulated in (BEL) for self-ascribing *perceptual* or *imagistic* states, as opposed to propositional attitudes, this could be applied by the mindreading system. And in fact, Carruthers does think that our access to our own perceptual states is transparent in this sense and not dependent on interpretation. Cf. Carruthers 2011: 72ff.)

Carruthers discusses a large body of empirical evidence and other considerations in favour of his proposal. For example, he argues that the global broadcast architecture is best suited to explain the possibility of a gradual development of the mind in incremental steps, where one system after another is aggregated through natural evolution. This also explains, he says, why nonsensory mental events such as judgments or decisions cannot be globally broadcast – the broadcast architecture was in place before such a redesign of the basic architecture could have been useful. He also discusses at length many cases in which subjects seem to confabulate what their own intentions, desires and even beliefs are. For example, hypnotised persons who receive an order, frequently, when woken up, comply with the order; say putting a book from the desk onto the shelf. But when asked why they do so, they explain that they dislike the disorder and decided to clean up or some such (cf. Wegner 2002). Carruthers interprets such cases as evidence for the view that self-attributions of propositional attitudes are based on unconscious self-interpretations. When we confabulate an intention that clearly is (or was?) not there, we interpret ourselves – how else should the self-attribution come about? Because we lack some relevant information our interpretation is erroneous. Since we are not aware that we are "just interpreting", it might well be that we always base our self-attributions on interpretations, even when they are true.

There is not enough space for me to discuss these arguments here. Instead, let me focus on one particular objection to Carruthers's theory of mindreading and on his reply to it, because they shed light on the relation between his account and transparency theories of self-knowledge such as Byrne's. As we have seen, Carruthers claims that the mindreading system does not have a general access to the subject's beliefs, intentions, decisions etc. Rather, just as all the other systems, it has to make do with the information it receives through global broadcasts of sensory information (and a limited amount of principles, data and so on specifically necessary for mindreading). But – this is the objection put forward by several commentators (cf. Currie and Sterelny 2000, Friedman and Petrashek 2009, Lurz 2009) – is it possible to interpret other people's minds without having a general access to one's own beliefs? It seems that we often need information about the world that is not perceptually present at the time of interpretation in order to attribute mental states to other people. We interpret them not only in the light of what we observe right now, but also in the light of what we believe about them and about the world in general.

Here is an example from Friedman and Petrashek: "Louise is an expert in British history, so she *knows* that the Battle of Hastings occurred in 1066" (Friedman and Petrashek 2009: 146). We attribute such knowledge (a propositional attitude) to Louise because we believe that the Battle of Hastings occurred in 1066, that Louise is an expert in British history and that experts in British history know such things. So reading the mind of Louise depends, in this case, on access to our own beliefs. It is imaginable that the three (supposed) facts in question are presented sensorily to the subject. For example, the subject might *read* about them, as you are now. However, while this *might* happen, it seems that no such sensory access is *necessary* for one to attribute the knowledge to Louise. It seems that the knowledge-attribution could proceed directly on the basis of our beliefs, without a sensory intermediary. If that is so, then we seem to have a counter example to Carruthers's theory.

In fact, it might be a general principle of mindreading, other things being equal, first to attribute the same beliefs to others as we have ourselves. If I take *p* to be true, then, without reasons to the contrary, I should attribute the belief that *p* to others as well. So in fact there is a rule such as (BEL) for attributing beliefs to *others*:

(BEL-3) If *p*, believe that Fred believes that *p*. (Byrne 2005: 96)

Although (BEL-3) is not, of course, as useful as (BEL) in producing true belief-ascriptions, it is still at least a good starting point for mindreaders.

Now, if these arguments are correct, then, contrary to what was said before, it seems that our mindreading system, or some other mechanism, does have nonsensory access to our own beliefs in the sense that it has access to what we believe. This means that it should not need to use the Rylean method for self-ascribing beliefs. If it can attribute the belief that the Battle of Hastings occurred in 1066 to Louise, reasoning from the (supposed) facts that it did occur on that date and that Louise is an expert in British history, then it should also be able to apply an epistemic rule such as Byrne's (BEL). It should be able, in other words, to reason from the (supposed) fact (i.e. from the belief) that Louise is an expert in British history directly to the belief that I *believe* that Louise is an expert in British history. All it needs for such reasoning is an epistemic rule such as (BEL).

Carruthers's reply to this objection is threefold. First, he concedes that rules such as (BEL) and (BEL-3) can be used by us, but he says that it is not the mindreading system that applies them and that they lead to merely verbal self-attributions of belief. Second, he concedes that the mindreading system can have access to all of the subject's beliefs, but only indirectly via the global workspace and operating in a slow and reflective, system 2-type of way. Third, he maintains that, in automatic or "online", system 1-type of operation, the mindreading system only has access to sensory information.

To focus on the second and third part of his reply first, it is clear that Carruthers does not think that the mindreading system ever uses rules such as (BEL) or (BEL-3). He maintains that it does not have any direct access to the subject's own beliefs. When mindreading happens in an automatic or "online" way (the third part of Carruthers's reply), the system mainly interprets occurrent sensory information and does not have access to the stored beliefs and other propositional attitudes of the subject. But there is a different, slower and more reflective way of operating in which the mindreading system can access all the subject's propositional attitudes. It can do so by posting queries in the common workspace of the mind. "The entire suite of consumer systems then gets to work, drawing inferences and reasoning in their normal way, accessing whichever of the subject's belief they normally would. The results are then posted back into the global workspace once more, [...] Here the entire process, collectively, has access to all of the agent's beliefs;" (Carruthers 2011: 238). In this reflective mode, the mindreading has access to all of the subject's beliefs, but only indirectly, via the global workspace. Since any information from the workspace is sensorial, it needs to be interpreted to yield information about the subject's propositional attitudes.

The first part of Carruthers's reply to the objection is more interesting in our context. For Carruthers concedes that we in effect use rules such as (BEL) and (BEL-3) to attribute beliefs. It is just that these rules are not implemented by the mindreading system, but by the executive and language-production systems and the result is not real self-knowledge (or knowledge others' beliefs). Rather, Carruthers seems to think that it is a purely verbal attribution that we can make in reply to a verbal question:

> If my task is to say which city someone believes to be the capital of the United Kingdom, for example, then I shall immediately answer, "London," without knowing anything further about that person. [...] the executive and language-production systems cooperate (and partly compete) with one another, searching the attributor's own memory and issuing the result in the form of a metarepresentational report – "I think/she thinks that P" – where the form of the report can be copied from the form of the initial question. (Carruthers 2011: 237)

This account of how we can come to make "a metarepresentational report" seems to be quite in line with transparency accounts such as Byrne's. The crucial difference is that Carruthers does not think that the report expresses self-knowledge (in case of having the form "I think that P") or knowledge of the beliefs of others (in case of having the form "She thinks that P"). Rather, the prefix (in the case of the first-personal report) is "a mere manner of speech or a matter of politeness (so as not to appear too confident or too definite)" (Carruthers 2011: 86).

What is curious about this position is that it does not seem to make a distinction between self and other-attributions of belief. What is the difference, in the speaker's mind, between "I believe that *p*" and "He believes that *p*", if the prefix is only a mere manner of speech? It seems that, according to Carruthers, there is none. Rather, it is only by interpreting our own verbal reports that we find out about whom we are talking. Yet more strangely, even if we say "He believes that *p*" we are not actually expressing a belief about some other person, but only the belief that *p*.

To conclude, where does this discussion leave transparency theories of self-knowledge? If we take Carruthers seriously, there is a place for transparent self-ascriptions of belief: It lies in our immediate and, presumably, unreflective verbal answers to questions about our beliefs. Here we directly access the first-order belief and merely verbally prefix it with "I believe that ...", thus applying Evans's procedure. But if Carruthers is right, the result is not self-knowledge, but a mere manner of speech. It is unlikely that this will satisfy transparency theorist such as Byrne. It also has the odd consequence that we often seem to talk about beliefs without knowing whether they are our own or someone else's.

## Literature

Byrne, Alex 2005 "Introspection", *Philosophical Topics* 33, 79-104.

Carruthers, Peter 2011 *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.

Currie Gregory and Sterelny, Kim 2000 "How to think about the modularity of mind-reading", *Philosophical Quarterly* 50, 145-160.

Evans, Gareth 1982 *The Varieties of Reference*, Oxford: Clarendon.

Friedman, Ori and Petrashek, Adam R. 2009 "Non-interpretive metacognition for true beliefs", *Behavioral and Brain Sciences* 32, 146-147.

Lurz, Robert W. 2009 "Feigning Introspective Blindness for Thought" *Behavioral and Brain Sciences* 32, 153-154.

Ryle, Gilbert 1949 *The Concept of Mind*, London: Hutchinson.

Wegner, Daniel M. 2002 *The Illusion of Conscious Will*, Boston: MIT Press.