

Transcendental Arguments in the Philosophy of Mind

Mathilde Byskov Jakobsen

Department of Philosophy, University College London

m.jakobsen@ucl.ac.uk

Abstract

In this paper I defend the use of transcendental arguments within the philosophy of mind. I argue that we can use transcendental arguments to establish certain necessary features and capacities of the mind, given that it is the mind of a certain kind of experience. This means that as well as formulating belief-directed transcendental arguments, we can also formulate truth-directed transcendental arguments as long as they are directed at certain features of the mind. I present an argument for the unity of consciousness to illustrate this form of transcendental argument. I argue that this is a truth-directed transcendental argument which is not subject to Stroud's objection to the use of such arguments outside of an idealist framework.

1. Introduction

My aim in this paper is to defend the use of transcendental arguments in the philosophy of mind. I will identify a certain kind of belief-directed transcendental arguments that can be of some limited use in the philosophy of mind. This use of transcendental arguments fits with the view that all transcendental arguments can ever hope to reach are conclusions about beliefs. However, the main claim which I will defend in this paper is that a certain kind of truth-directed transcendental argument can also be of use in the philosophy of mind. I will defend this claim primarily by considering the methodology of transcendental arguments directed at the mind, but also by developing an example of a self-directed transcendental argument. I call this type of argument a "self-directed transcendental argument" following Quassim Cassam, as it is a name which reflects its purported subject matter. My paper has three main parts. First, I will consider the methodology of transcendental arguments in general and present Barry Stroud's well-known objection to them. Second, I will present two kinds of self-directed transcendental arguments of use in the philosophy of mind, which are not subject to this objection. The first of these is a kind of belief-directed transcendental argument. The second kind is a kind of truth-directed transcendental argument. The possibility of formulating truth-directed transcendental argument of this kind is exclusive to the philosophy of mind, and it is dependent on an assumption about the nature of mind which is broadly functionalist. In the third part of the paper I will further evaluate this strategy by considering an example of a self-directed transcendental argument. This argument will be developed from some of Kant's claims about transcendental apperception in the 'Transcendental Deduction'. I conclude that this methodology can identify significant constraints on the form of the mental

operations that our minds must implement, given that our experience has the character that it does, and so tell us something about the mind – or the self.

I

2. What are transcendental arguments?

Transcendental arguments argue from premises about experience to a conclusion about reality. Typically, they start with a premise which states that we have experience of a certain kind. The second premise determines that this form of experience is conditional upon some unobvious and philosophically interesting claim about reality. Which we can then conclude must hold. In one traditional use of such arguments, the aim of the argument is to refute the philosophical sceptic. In this traditional use the first premise is selected in order to be immune from Cartesian scepticism about the external world. This premise describes experience in a way which cannot be doubted from the first person perspective. However, this aspect is optional: transcendental arguments are not, primarily, intended to refute the philosophical sceptic even if they have, traditionally, been used in this way. Rather they are concerned essentially with identifying necessary conditions for experience, hence the force of the description “transcendental”. This means that instead of having the following form:

Necessarily P
Necessarily, (if P then Q)
Therefore, necessarily Q

The transcendental arguments I will discuss have this logical form:

P
Necessarily, (if P then Q)
Therefore, Q

If the premise we start with is only contingent, - if it is taken as a matter of fact that we have a certain kind of experience, then the conclusion reached is not necessary. And if we cannot establish the first premise with certainty, then the conclusion reached is only conditional. This means that the argument cannot be used to refute philosophical scepticism. On the other hand it is not dependent on the claim that the experience we have is in some sense necessary.¹

An example of a transcendental argument can be found in Kant’s refutation of idealism (B275):

I have experience of myself as determined in time;
Time determination is dependent on the experience of something persistent, and this cannot be merely a representation in me;

¹ Harrison (1982) p. 216.

Therefore, there must be something persistent outside me of which I have direct consciousness by which I can judge the passing of time.

3. Stroud's Objection

In a well-known article from 1968 Stroud objected to this entire class of arguments. His main objection is that transcendental arguments cannot draw conclusions about reality, because it is always open to an objector to insist that it is enough that the world appears to us a certain way or that we believe it to be a certain way in order for us to have experience – it is therefore not necessary that the world is in fact in this way. We can apply this objection to Kant's argument above. It draws a conclusion about reality – that there is in fact something persisting outside of me of which I can be conscious, but, the objection goes, it is in fact enough if it just appears to me as if there is something persistent outside of me and that I believe that it is independent of me, in order for me to have experience of myself as determined in time.² If this is true, then the transcendental argument has reach a conclusion about some belief or experience that a thinker must hold, given that the thinker holds other beliefs or have other kinds of experiences. It then does not in fact argue from a premise about experience to a conclusion about reality. TA's are therefore only belief-directed, rather than truth-directed and do not establish objective conclusions about how reality must be.

This objection, however, does not hold against the use of transcendental arguments within the framework of transcendental idealism, as Stroud has also made clear.³ Such a framework secures the bridge between experience and reality, because reality, as phenomena, is dependent on our experience of it. Therefore, in drawing conclusions about necessary features of our experience using transcendental arguments one is drawing conclusions about phenomenal reality. As Kant puts it "The conditions of the possibility of experience in general are likewise conditions of the possibility of the objects of experience, and (...) [for this reason they have objective validity in a synthetic a priori judgement]" (B197). This means that within the framework of transcendental idealism we can formulate *truth*-directed transcendental arguments.

II

4. Psychological belief-directed, self-directed transcendental arguments

We can concede to Stroud that transcendental arguments must be interpreted as *belief-directed* if they are to be valid outside the framework of transcendental idealism, and still preserve the usefulness of such arguments. We can do that by re-deploying these arguments, to be not about the world,

² Stroud (1968).

³ Stroud (1999) p. 160.

but about how we must take the world to be – not about facts but about beliefs.

So far I have discussed transcendental arguments as directed at either the world or a belief about the world. However, we can also direct transcendental arguments at the mind or the self or at a belief about the mind or the self. I will call these arguments self-directed arguments, following Cassam.⁴ Self-directed transcendental arguments argue from premises about experience to a conclusion about a belief or a fact about a feature or a capacity of the mind or the self which is necessary for the experience in question.

We can produce the following grid of the two kinds of transcendental arguments so far identified:

World or Self/Mind	1: Truths
	2a: Beliefs
	2b: Experiences

The lesson learned from Stroud is that in a non-idealist framework we can only formulate transcendental arguments of the second, belief-directed, kind.

Cassam's argument in *Self and World* is an example of a belief-directed, self-directed transcendental argument. The claim of the transcendental argument is that it is necessary that a subject of self-conscious experience has a psychological belief with a certain content: oneself qua subject as a physical object. Cassam develops both a belief and an experience version of this argument. He develop the experience version of the argument because the belief-version is subject to the objection that people who believe that they are Cartesian souls are still able to ascribe their representations to themselves is solved by saying that they in fact still have an experience or an intuition of themselves as physical objects, however confused they are in their beliefs.

Belief-directed transcendental arguments clearly have a use in the philosophy of mind. Arguments of this type can show the necessity of a certain psychological belief that a subject must have in order to have experience. I call these kinds of transcendental arguments 'psychological belief-directed arguments' for this reason. Such arguments can show that a subject, who has experience of a certain kind, must have some actual belief with a specific content. The belief is psychologically necessary for the possibility of experience. However, all such arguments are subject to what Cassam calls the 'misconception' objection (*Self and World*). The misconception objection turns on the fact that if a transcendental argument can show that some belief is necessary for experience, then it should not be possible to have a misconception about the subject matter of this belief. If a belief-directed transcendental argument for instance shows that you cannot have experience of yourself in time without believing that there are objects

⁴ Cassam (1999).

outside of you, then the existence of sincere idealists, who experience themselves as determined in time, is impossible.

Psychological belief-directed, self-directed transcendental arguments can only be used to show that some belief, which we in fact all hold, is necessary. This obviously does not make such argument useless: if successful, the transcendental argument will tell us that it is not a contingent matter that we all hold the same belief. However, it does limit the subject area of such arguments radically.

However, the use of transcendental arguments in the philosophy of mind is not limited to establishing the status of certain beliefs that we all hold. Rather, philosophy of mind is an area where we can formulate *truth*-directed transcendental arguments, or so I will argue. In order to argue for this claim, I will first identify an alternative way in which transcendental arguments can be belief-directed and met Stroud's objection. Second, I will argue that in the area of the philosophy of mind some such belief-directed, self-directed transcendental arguments are in fact also truth-directed transcendental argument. If this is right, then transcendental arguments have the potential to draw conclusions about facts about the mind or the self, not just about beliefs about the mind or the self.

5. Theoretical belief-directed, self-directed transcendental arguments

We introduced psychological belief-directed transcendental argument as a reaction to Stroud's objection to truth-directed transcendental arguments. Psychological belief-directed transcendental arguments met Stroud's objection by making their aim less ambitious. Psychological belief-directed arguments aim to draw the conclusion that a subject of experience must have a certain belief, not that a certain fact must be true of the world or the subject.

However, that is not the only way in which we can make our arguments less ambitious and so avoid the problem Stroud points out. We can re-deploy transcendental arguments to reach conclusions about our conceptual scheme, instead of about the world. This solution can be used on Strawson's arguments in *Individuals*. On this view transcendental arguments can aim to show us that a certain belief is part of our conceptual scheme. This designates a different way in which transcendental arguments can be belief-directed than the way discussed above. Transcendental arguments can aim to show conceptual connections between the way in which we, as theorists, think of experience, and the way in which we, given that, have to use related concepts. This importantly does not place any psychological constraints on the beliefs of individual people. I call this second sense of a belief-directed transcendental argument, a *theoretical belief-directed argument*, because it shows the theoretical necessity of a belief, rather than the psychological necessity of the belief. A transcendental argument could for instance attempt to establish that given our conceptual scheme it is necessary that we take objects to be in space and time. The conclusion, that objects are in space and time, is a belief rather than a fact or truths about the world. However, the argument does not establish that single individuals must hold this belief to have experience. This distinguishes theoretical belief-directed transcendental arguments from psychological belief-directed transcendental arguments.

Cassam claims that all belief-directed transcendental arguments are subject to the misconception objection [*Self and World*]. However, in fact the misconception objection is only effective against psychological belief-directed arguments. Sincere idealists are not a problem for the transcendental philosopher who formulates theoretical belief-directed arguments. She is committed only to saying that a certain belief is (theoretically) required given our conceptual scheme not that it is psychologically required, and so allows that people can be confused and still have experience. A theoretical belief-directed version of Cassam’s argument for the necessity of having the belief that one is a physical object to have self-conscious experience would not be subject to the misconception objection. It would just show that we have good philosophical reasons given our concepts of experience and subjects to think that subjects have to be physical objects, without demanding that the subject must actually believe itself to be one. This argument meets Stroud’s objection. It does that because it allows that, though it is part of our conceptual scheme that we take ourselves to be physical objects, we cannot prove that we really are physical objects. It allows for the possibility that the world may in fact be very different from the way we have to take it to be.

We can add this last type of transcendental argument to the grid from before, so that it now looks like this:

World or Self/Mind	1: Truths	
	2a: Beliefs	} psychological
	2b: Experiences	
	3: Beliefs	} theoretical

Theoretical belief-directed transcendental arguments are not very interesting, when directed at the world in a non-idealist framework. Outside of transcendental idealism conditions of cognition are not also conditions of objects. The way in which we have to think of things does not tell us anything about how they are. In the case of world-directed transcendental arguments, we can for instance attempt to show that the belief in causation plays a central role in our conceptual scheme, so that in a reflection on the possibility of experience, we have to think of the world as causally ordered. However, the question will always arise whether independent reality actually corresponds to this necessary way of conceiving of it. The fact that our conceptual scheme demands that we make sense of the world only as causally ordered does not tell us what we want to know, namely whether the world is in fact causally ordered or not.

6. The nature of mind

However, my suggestion is that in the case of at least some of the transcendental arguments we can direct at the mind, we do not find the same gap between how we must think of things and how they really are. Say that we have established that we must necessarily interpret or conceive of the mind as having certain features and characteristics in order to understand it as a mind which has experience of a certain kind. Now, we cannot allow that it

may in fact not really have these features at all – whilst still being interpretable as a mind of this kind of experience.

In cases where our argument shows that we must necessarily think of the mind as realising certain *functions or features related to functions* there is no sense in driving a wedge in between the claim that the mind must be interpreted or conceived as having these features and capacities, and the claim that the mind really realises these features and capacities. The picture of the mind required to support this view is broadly functionalist in that a mind is defined by its functions.⁵ If something is defined by its function, and it performs these functions, then there is no question as to whether its real nature fits our functional picture of it.

This assumption about the mind opens up the possibility that *if* we can formulate transcendental arguments about features of the mind necessary for it to perform its functions, then we can consider them as identifying aspects of the mind and not just of our conception of the mind.

This assumption about the mind must not be confused with the claim that the mind is transparent to itself either in the sense that we know infallibly the features of our minds or the content of our mental states by direct introspection. It is not a claim about people's psychological relation to their empirical minds.

It also does not suggest that we can find out everything there is to know about the mind by the use of a transcendental reflection on the possibility of experience. First, there are many empirical facts about the mind which require empirical methods. Second, transcendental philosophy cannot tell us how a certain necessary feature or capacity is realised in the mind – if there is more than one way in which this could be done. A transcendental philosophy of mind uses self-directed transcendental arguments to ascribe features to the mind (necessary conditionally dependent on that the mind in question is a mind of a certain kind of experience). However, while it tells us that the mind must realise these features, it does not tell us how it realises them. Just like other kinds of functionalism, transcendental philosophy of mind is ontologically uncommitted. It can tell us nothing about either the substrata of mind or about the actual realisation of the very general constraints we can place on the mind.

This suggests that the philosophy of mind allows for transcendental arguments to be used to establish facts about the mind and not just beliefs about the mind.

III

7. An example

To take a concrete example of this strategy, consider Kant's account of transcendental apperception in the B-edition of the 'Transcendental

⁵ It is only 'broadly' functionalist as it is not committed to any of the more specific requirements of contemporary functionalism.

Deduction'. A self-directed transcendental argument can be developed out of some of these considerations. The starting point of the B-Deduction is the observation that experience is made up of complex representations, where both the individual elements of a complex representation and different complex representations are synthesised together to form one experience. The next step is to say what this requires in the subject of such experience. The answer for Kant is the unity of transcendental apperception:

It must be possible for the 'I think' to accompany all my representations; because otherwise something would be represented in me which could not be thought at all and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me. (B132).

This quote is saying that all conscious parts of one experience (all representation that are not nothing to me) must be self-ascribable to one 'I'. The idea that *a bundle of representations are all in some sense mine* corresponds to one understanding of the unity of consciousness in contemporary philosophy of mind. We have to distinguish this understanding of the unity of consciousness from one that refers to a feature of our experiences themselves. The term 'the unity of consciousness' can be used to refer to a phenomenal feature of experience: the fact that our experiences are unified in the sense that it feels different to experience *a* together with *b*, than it does to experience *a* alone (and that it is the experience of *a*, which is different).⁶ Unity of consciousness, as I am interested in it, is not a claim about the feel of experience. If it was the argument would be trivial as the claim that experience is unified (at least partly) is already involved in the claim that it contains complex representations. Saying that a bundle of conscious states must all be ascribable to the same 'I', saying that a bundle of representations all in some sense belong to me, is not a claim about experience but about the mind. So the possibility of complex representations requires that the parts of this representation are in some sense all mine – which is equivalent to the claim that they belong to one unity of consciousness. I will suggest three ways in which we can understand the unity of consciousness understood in this way.

One way of understanding what it means to be self-ascribable to one 'I' is via a causal relation to a body – a bundle of experiences are part of a unity of consciousness if they are all causally dependent on the same body. We can reject this view because there are no good reasons to think that there is a one-one relation between a body and a unity of consciousness – and split brain and multiple personality disorder cases in fact give us reasons to think that this is not necessarily the case. The conceivability of a unity of consciousness supported by more than one body – perhaps by a succession of bodies as in Kant's billiard ball example also supports this point. The fact

⁶ Dainton and Chalmers and Bayne take the unity of consciousness to express a phenomenal quality about our experiences themselves at a time. For Dainton the unity of consciousness is a phenomenal characteristic that lies in the co-conscious experiences themselves, for Chalmers and Bayne it is the existence of a state, whose phenomenal character subsumes the phenomenal characters of all the states the subject is in at that time.

that a bundle of representations all belong to a body in being causally dependent on it does not mean that they can all be self-ascribed together. And the possibility that a bundle of representations are causally dependent on two bodies does not rule out that they could be self-ascribed to one subject together. And it is this latter sense of belonging we need to make complex representations possible. This shows that just because something is true of the subject it is not necessarily true for the subject - it is not enough that it is true that the parts of a representation are mine if I don't take them to be mine.

Another option is to understand the unity of consciousness as a claim about actual self-ascription. Under this suggestion a number of conscious states are part of a unity of consciousness if there exist in the subject a state of the form 'I think that...' which subsumes all of them. However, the problem with basing unity of consciousness on actual self-ascription is that it limits it to include only states where such a higher order state exists, which seems too restrictive. It is after all supposed to be necessary for the possibility of complex representations.

We therefore have the further option of basing the unity of consciousness on *possible* self-ascription, which avoids this problem. According to this suggestion a bundle of experiences are parts of a unity of consciousness if they *can* be self-ascribed together under one 'I think'. Some people object to the use of dispositions or possibilities in an account of the unity of consciousness because it is hard to make the claim substantial. In situations where a subject at a time has a whole range of experiences, only some of which are ever actually taken together, how do we know that it could actually have done so to all of them? However, the solution cannot be to avoid talk of possibilities in favour of actualities. Accounting for the unity of consciousness in terms of a disposition or a capacity for self-ascription is still the best option, in my view, because it captures the sense of experiences belong to me, that we need, without being unnecessarily restrictive.

If experience contains complex representations, then the subject of experience realises a unity of consciousness – that is, the parts that make up the complex representations can all be self-ascribe under one 'I think'. Kant illustrates the necessity of the unity of consciousness with the example of the consciousness of a verse of a song – in order for there to be *a consciousness of a verse*, it is necessary that all the parts of the verse is taken by the subject as all its experiences and as belonging together. If a conscious complex representation is to count as such, then the subject must be able to take its part as a whole and self-ascribe it as one. There is no need for the subject to actually so self-ascribe it, which shows that it is the possibility not the actuality of self-ascription which is important. So if experience contains complex representations, then the unity of consciousness is necessary for experience.

8. The status of the conclusion

In this argument we go from an observation of the kind of experience we have to a conclusion about a necessary feature of the mind that has such experience, which in this case is the unity of consciousness. This conclusion of the argument is not that the subject must believe in the unity of consciousness in order to have experience. If the unity of consciousness is necessary for the possibility of experience containing conscious complex

representations, then it is neither necessary nor sufficient that the subject *believes* that it realise a unity of consciousness, it must actually do so. The argument does make a claim our conceptual scheme. The conclusions we draw about the mind on the basis of self-directed transcendental arguments link ways in which we think of experience with ways in which we, given this, must think of the mind. The argument identifies a conceptual connection between thinking of experience as unified and thinking of the subject of these experiences as realising a unity of consciousness. Describing this connection as conceptual does not reduce it to the trivially linguistic. It involves a reflection on experience, and a reflection on how we must think of the experiencer given that it is the subject of this kind of experience. However, the conclusion is not restricted to a claim about how we must think of the mind. In the case of functionally defined aspects of the mind, if we can show that we must necessarily think of the these features in a certain way, given that it is the mind of a certain kind of experience, then this identifies a real aspect of the mind.

It seems to me plausible to defend this general claim in the particular case of the argument about the necessity of the unity of consciousness.

This particular use of a self-directed transcendental argument, then, is not subject to Stroud's general methodological strictures: It is not restricted to showing only the necessity of a psychological belief that we all hold and it is not restricted to showing only something about our general conceptual scheme – if we take this to mean that it has not also been established as a feature of reality. The argument is therefore truth-directed.

I conclude that the philosophy of mind allows for the use of transcendental arguments to determine the features of the mind that are conditionally necessary for our experience to take the form that it does.

Bibliography

Brook, Andrew: *Kant and the Mind*. Cambridge University Press 1994.

Cassam, Q: *Self and World*. Oxford University Press 1997.

Cassam, Q: 'Self-directed Transcendental Arguments' in Stern, R: *Transcendental Arguments*. Oxford University Press 1999.

Chalmers, D. and Bayne, T: 'What is the Unity of Consciousness'. Here from Chalmers' webpage, but published in A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, Dissociation*. Oxford University Press 2003.

Dainton, B: *Stream of Consciousness*. Routledge 2000.

Gardner, S: *Kant and the Critique of Pure Reason*. Routledge 1999.

Harrison, R: 'Transcendental Arguments and Idealism' in Vesey, G: *Idealism, Past and Present*. Cambridge University Press 1982.

Kant, I: *Critique of Pure Reason*. (Translated by Kemp Smith) The Macmillan Press 1929 (second edition 1933).

Rosenthal, D: 'Unity of Consciousness and the Self' Proceedings of the Aristotelian Society 2003.

Strawson, P. F: *Individuals*. Routledge 1959.

Stroud, B: 'Transcendental Arguments' *The Journal of Philosophy* vol. LXV, no. 9. 1968.

Stroud, B: 'The Goal of Transcendental Arguments' in Stern, R: *Transcendental Arguments*. Oxford University Press 1999.

Stroud, B: 'The Synthetic A Priori in Strawson's Kantianism' in Glock, H-J: *Strawson and Kant*. Oxford University Press 2003.