



**MARIA MATUSZKIEWICZ**  
UNIVERSITY OF WARSAW

## **KNOWLEDGE ABOUT OUR EXPERIENCE AND DISTINGUISHING BETWEEN POSSIBILITIES**

In his John Locke lectures delivered at the University of Oxford and published as a book *Our Knowledge of the Internal World* (2008), Robert Stalnaker characterizes the difference between two opposing philosophical perspectives: the externalist and the internalist one in terms of their starting points. For an internalist, it is mental contents accessible through introspection that are the foundation of knowledge. The philosophical question an internalist asks is *how can our knowledge reach beyond the contents of our mental states?* For an externalist, it is the external world with the objects, properties, and relations within it that are the starting point. Among these objects there are creatures who have thoughts and experiences. The question is: *how can these objects – human beings – have thoughts, which are about the world and about themselves?*

Stalnaker argues that puzzles concerning knowledge about our experience originate in the conflation of the two perspectives. As a remedy, he proposes a more thorough form of externalism. Externalism conceived in such a way consists in not only the claim that the contents of our mental states are determined by external facts (including both natural facts and the social environment), but also in the claim that contents are essentially ascribed. This type of externalism is linked to contextualism: contents are always ascribed in a particular context and there is no single correct characterization of our mental state, independent of the context of ascription. Stalnaker argues that providing a solution to some philosophical problems, problems of intentionality and of knowledge included, requires a shift from the perspective of a subject to the perspective of a theorist.

The main problem that the book addresses concerns our epistemic relation to our experience and the relation between experience and knowledge. Discussing Frank Jackson's knowledge argument and a solution to the puzzle it describes, Stalnaker argues that, contrary to what the empiricist tradition takes for granted, our knowledge about our experience is no more direct than our knowledge about external objects. Stalnaker's solution to the knowledge argument is based on the analogy between phenomenal and self-locating knowledge. Both are accessible only from a particular perspective. The most complex parts of the argument presented in the book concern the relation between the knowledge we can have only from a certain perspective and the objective knowledge.

### **The knowledge argument and the three strategies**

Stalnaker starts with a discussion of the knowledge argument and three different strategies to avoid its conclusion.

The knowledge argument is based on two premises: (1) Mary knows all the facts of the type F. (2) Mary doesn't know the fact that  $p$ . From the two premises it follows that (3) The fact that  $p$  is not of the type F. The argument leads to the conclusion that beyond physical facts (or communicable facts) there exist facts of a different kind, phenomenal facts. Stalnaker discusses three strategies to resist the argument. The Fregean strategy adopts a more fine-grained notion of information than that of discriminating possibilities. David Lewis claims that what Mary acquires is not new information, but a new ability. John Perry's solution draws on an analogy between phenomenal and self-locating knowledge. All three strategies attempt to rebut the argument by ruling out the possibility that what Mary lacks is certain information, understood in terms of distinguishing between possibilities. It is this possibility that Stalnaker urges us to recognize.

#### ***i. The Fregean strategy***

A proponent of the Fregean solution claims that Mary does not learn a new fact, but she learns the same fact in a new way. The solution requires modes of presentation or senses which individuate thoughts in a more fine-grained way than in terms of their truth conditions. Before leaving the black and white room, Mary has knowledge about certain mental state under a functional or neurophysiological mode of presentation. As a result, she knows the same fact under a visual mode of presentation.

Stalnaker rejects this solution, because, all in all, it does not avoid the conclusion that Mary's lack of knowledge stems from her inability to eliminate possibilities, and at the same time it does not provide a good account of these possibilities. The argument takes the following form: if materialism is true, both of these concepts (modes of presentations) necessarily pick out the same object. We might have both concepts and yet not know that they apply to the same object: no *a priori* reasoning leads from the one to the other. Hence, we might conceive of a situation in which a certain object is  $\phi$ , without being  $\psi$ . We are forced to accept that the situation is epistemically possible, although it is not metaphysically possible. Stalnaker, however, rejects the idea that metaphysical possibilities are a proper subset of a broader class of possibilities, which include conceptual possibilities. His objection is that we have no conception of a merely conceptual possibility. Usually what is merely conceivable is defined in terms of what one may have a clear conception of. Stalnaker argues that we cannot have a clear conception of an impossibility. In cases of necessary *a posteriori* truths, one way of explaining modal illusions is to redescribe them: for example a situation where one thinks that water is not H<sub>2</sub>O can be redescribed as a situation where one thinks of a substance phenomenally alike water that it is not H<sub>2</sub>O. Phenomenal experiences do not – one might argue in Kripke's line of thought – allow for such a redescription. Let's suppose that having a red sensation is identical to a functional state F. Yet, I can think that it is not. The possibility that I conceive of – one would argue – cannot be adequately redescribed as a possibility that some experience, other than the sensation of red, is not a functional state F. It is because the phenomenal property (the property of being experienced as seeing red) is essential to seeing red. Stalnaker presents this view just to reject it. As the objection is a step in his argument against the Fregean solution, one might worry whether – since Stalnaker himself rejects this objection – he does not dismiss this strategy too fast.

**ii. *The ability hypothesis***

Lewis (1988) rejected the assumption (2) that what Mary lacks is knowledge of a certain fact. What she acquires after leaving her room is a new ability, not knowledge. This ability is not of a cognitive kind. Cognitive abilities enable us to distinguish between possibilities. Lewis argues that neither Mary's situation before her release nor afterwards

can be described in terms of an ability to distinguish between possibilities. Lewis argues that before having an experience one cannot represent different possibilities. Before her release, Mary could not represent what it is to see red or what it is to see green and hence – according to Lewis – these possibilities cannot be used to characterize her mental state. After her release she can only think that *it is like that to see red* and again she cannot distinguish between different possibilities. Stalnaker points out that *post factum* Mary can represent different possibilities that she could not represent before and that these possibilities can be used to characterize her past cognitive limitations. What is important is that possibilities play an external role in characterizing the thinker’s mental states: it is the theorist who uses them in order to ascribe mental contents. Stalnaker argues that the ability that Mary acquires is a cognitive ability. “While it may be right, as the ability hypothesis claims, that Mary does not necessarily acquire information merely by having color experience, it seems that she does acquire an ability to make distinctions between possibilities that she could not distinguish before, and a proper account of these abilities requires an account of the distinctions between the possibilities” (Stalnaker 2008:37). The quote is key to understanding Stalnaker’s view on the relation between the knowledge about our experience and the knowledge about the world.

### ***iii. The self-locating analogy***

John Perry’s strategy is, according to Stalnaker, the most promising one. It draws on the analogy between self-locating and phenomenal knowledge. The analogy suggests how to avoid the conclusion of the argument. While we might know all facts of type F and yet still lack some self-locating knowledge, very few philosophers would conclude that there are self-locating facts ‘over and above’ physical facts.

Perry’s solution proposes that beliefs and utterances have more than one type of content: aside from subject-matter content, they also include reflexive contents (Perry, 2001). Subject-matter contents are conditions that the world has to satisfy for the belief or the utterance to be true. Reflexive contents are conditions not only on the world but also on the belief or the utterance itself. When John says “I am happy”, the subject matter content of his utterance is *that John is happy*. Its reflexive content is *that the person having this thought token is happy*. In some

situations, all we learn when acquiring a self-locating belief is a reflexive type of content. For example, I know that the meeting starts at noon. When suddenly I realize that the meeting starts now, all I learn is a reflexive content: that the time of this thought token is noon. What I learn does not commit me to the idea that there are some self-locating facts over and above physical facts. Perry adopts this strategy to solve the puzzle about Mary: what Mary learns upon her release is a belief with a different reflexive content. Before her release Mary knew (1) *QR is what it is like to see red* (*QR* being a functional concept of seeing red). What she learns upon her release is (2) *this<sub>RED</sub> is what it is like to see red* (*this<sub>RED</sub>* being a visual concept of seeing red). The two thoughts have the same subject matter content, but they differ in their reflexive content: “(1) is true iff the origin of Mary’s *QR* concept, the concept involved in (1), is the subjective character of the experience of seeing red”. (2), on the other hand, “is true iff the act of inner attention to which it is attached is of the subjective character of the experience of seeing red” (Perry 2001: 147-148).

While Stalnaker accepts the general intuition behind the notion of reflexive content and agrees that an adequate theory of beliefs requires an account of how we can represent the perspective from which we perceive the world, he criticizes the way Perry has introduced the notion. First, he argues that Perry confuses the means of a representation with its contents. While the distinction between subject matter and reflexive content applies to utterances, its application to beliefs is problematic (Stalnaker 2008:39-40). Utterances are different than beliefs in that we might individuate them also in terms of occurrences of certain patterns of sounds, not only in terms of their content. With regard to beliefs, we cannot individuate them in any other way than in terms of their content<sup>1</sup>, that is, in terms of the proposition they express. Thus, when we speak about different types of contents regarding beliefs, we must assume a vehicle of content: language of thoughts or inner symbols representing the content. Stalnaker insists that instead of distinguishing between different types of contents we should model all the relevant aspects of

---

<sup>1</sup> We might of course individuate a thought by referring to it as *Mary's favorite thought*. We may, however, still ask *what is her favorite thought?* In order to know whether Mary's favorite thought is the same as John's we need to know their contents.

content (including the thinker's perspective) in terms of distinguishing between possibilities.

It is worth noting that Stalnaker is generally critical with respect to theories which attempt to account for the intentional character of our thoughts by postulating concepts, mental files, and other inner vehicles of contents. His main charge is that these theories confuse intentional with non-intentional descriptions and they merely pretend to explain intentionality. For example, the mental file metaphor explains the difference between two beliefs with the same subject matter content, in terms of the difference in the mental files involved, which 'store' these contents. Stalnaker reminds us however that what these files supposedly consist of are not propositions, but certain physical objects, whose intentional properties still require explanation (Stalnaker 2008: 40).

Stalnaker also points out that reflexive contents as such do not explain Jackson's puzzle. He evokes Nida-Rümelin's thought experiment to demonstrate that the analogy between what we learn when we acquire a self-locating belief and what we learn when we have a phenomenal experience is flawed. In Nida-Rümelin's scenario, Mary's cognitive achievement is divided into two steps (Nida-Rümelin, 1995). Upon her release, Mary is first transported to a room covered with a multi-colored abstract wallpaper. Mary experiences colors for the first time, but she cannot connect her sensations to the concepts she has had before. It is only at the second step that Mary learns which color is which. It is then that she acquires information which is analogous to a self-locating belief. Thus, the analogy – Stalnaker concludes – cannot explain what she learns at the first step.

Stalnaker discusses yet another – although in his own view apparent – difficulty with the analogy. Many philosophers would say that there is an asymmetry between the two types of knowledge: while the information one learns when he acquires a self-locating belief is a contingent one (what I learn from you saying "I am Smith", can be redescribed as: "a person looking such-and-such is Smith"), the information one learns by having an experience is not contingent in that sense. In the first case, the argument goes, we are allowed to interpret the demonstrative reference descriptively, as a non-rigid designator. In the second case, "*this* is what it is like to see red", doing so would amount to saying that in a different possible world some other experience would play the role of seeing red. Many philosophers reject such a possibility.

Stalnaker in his critique of direct knowledge aims to show that they are wrong.

### **Stalnaker's account of self-locating beliefs**

The analogy is helpful, according to Stalnaker, on the condition that we have an adequate theory of self-locating beliefs. The account Stalnaker proposes differs from Perry's and Lewis' (Lewis, 1979) in that he claims that we can explain the special nature of *de se* beliefs in terms of their content without introducing a special type of contents (reflexive contents) or without altogether modifying the notion of content (Lewis' centered worlds). The solution rests on the assumption that a lack of a self-locating knowledge always amounts to ignorance with respect to which of the worlds is the actual one. For the solution to work, it must provide such a mode of a transworld identification of the thinker, which excludes the possibility of the thinker's not knowing that the one who is thus identified is he himself. The mode of identification which satisfies this condition is by reference to one's occurrent thought token (Stalnaker 2008: 61).

Let's recall Lewis' example with two gods (Lewis, 1979): one god lives on the highest mountain and throws manna, the other god lives on the coldest mountain and throws thunderbolts. The gods are omniscient in the sense that they have all propositional knowledge. What they don't know is which one is which. According to the account proposed by Stalnaker, what each of the two gods doesn't know is which of the two worlds is the actual one: the world in which a person having *this* thought is a god living on the highest mountain or the world, in which a person having *this* thought is a god living on the coldest mountain. Stalnaker calls his solution a *haecceitistic* one, claiming that the worlds thus distinguished are qualitatively indiscernible (Stalnaker 2008: 58-59). This may give rise to three concerns: (1) whether the identification by reference to an occurrent thought token is really immune to the error through misidentification (that is, whether it is such that the thinker cannot be unaware that it is him who is thus identified) (2) that the solution commits us to existence of possibilities accessible only from a first-person perspective; what is the relation between the self-locating knowledge and the objective knowledge? (3) are the differences between the two worlds really merely *haecceitistic*, and do we have good reasons to think that some possibilities do not differ qualitatively?

### **Essentially indexical information and the relation between it and the objective information**

Stalnaker emphasizes that our theory of beliefs requires a notion of informational content which can be separated from its relation with the thinker whose knowledge it represents. We may have complete objective knowledge about the world, yet lack some self-locating information. Is then the self-locating information something “over and above” the objective information?

Not many philosophers are likely to draw this conclusion, unlike in the case of phenomenal knowledge. The analogy between the two kinds of knowledge is that in both cases the epistemic situation of the subject is represented by possibilities which can be distinguished only from a particular perspective. We need to explain the relation between the self-locating knowledge and the objective knowledge.

A notion important to Stalnaker’s view is the notion of *essentially indexical information*. Essentially indexical information consists of “distinctions between the possibilities (the ways the world might be) that can be represented only from a certain perspective, but that once represented, can be abstracted from the perspective” (Stalnaker 2008:78). We might explain the notion using one of the examples discussed by Stalnaker. Sleeping Beauty (the heroine of Adam Elga’s puzzle), before being put to sleep learns that she will be woken up once (on Monday) or twice (both on Monday and Tuesday) depending on the result of a coin toss (Elga, 2000).

On Monday (and Tuesday, should she be awakened then), Sleeping Beauty was able to distinguish between a world in which, as she would put it then, *today* is Monday, and a different world in which *today* is Tuesday. On Sunday she was unable to distinguish between these two possible worlds, since in both of them an event of the same kind occurred on both Monday and Tuesday. To distinguish one from the other, one had to be there, or alternatively, to remember later having been there: one had to be in a position to refer uniquely to *that* particular time that Sleeping Beauty was awakened. But even on Sunday, Beauty was able to *describe* the distinction she was unable to make (Stalnaker 2008:78).

The notion of essentially indexical information is best understood in the light of Stalnaker’s theory of communication. Communication always takes place in a context (which includes beliefs held by the participants in the conversation) and results in a change of context. The context is

best represented as a set of possibilities (possible worlds). While communicating, the participants change the context by adding new information i.e. by excluding some possibilities. Sometimes knowledge of the relation between the utterance and the context is important to determine what information is being communicated. It is, however, not itself part of that information. For example, when I tell my friends “I live in Warsaw”, they can extract the information from the context in which I communicated it. On the other hand, when I introduce myself by telling you my name, the information I thus convey cannot be extracted from the context of utterance, because “there is not a piece of information that is the content of what I told you that you can simply add to your stock of beliefs about the objective world”. As Stalnaker explains, “the point about essentially contextual information is that sometimes the content of what is expressed or believed in is not detachable from the context in which it is expressed or believed” (Stalnaker, 2008: 81).

Whether one represents the thinker’s beliefs by means of locally-distinguishable possibilities or by means of non-local possibilities depends on his (the theorist’s) goal. Stalnaker illustrates this kind of context-dependence of belief ascriptions with the following example (Stalnaker 2008: 83-84): imagine Rudolf Lingens, the famous amnesiac from the Stanford Library. His two colleagues Daniels and O’Leary, who don’t know his true identity, call him “Nathan”. One day they see a crowd of journalists gathering in front of the library. O’Leary asks his friend: “Do you know who it is?” pointing at a man surrounded by journalists. “Yes” – Daniels replies – “it is our famous amnesiac friend, Nathan”. In this context – as Stalnaker argues – we can represent Daniels’ beliefs straightforwardly as beliefs about Lingens: the possibilities he eliminates are those in which it is someone other than Lingens whom they see. On the other hand, if one of the journalists approaches Daniels and asks him “Do you know who it is?” pointing at Lingens, the theorist would like to emphasize Daniels’ ignorance with respect to his friend’s true identity and he would represent Daniels’ epistemic situation using locally-distinguishable possibilities: *the man who is there is X, Y, Z etc.* What is local about this characterization of Daniels’ beliefs is that it is only relative to this context that we cannot describe Daniels’ beliefs in terms of Lingens himself and that the possibilities that we use to characterize his beliefs are distinguishable only within that particular context.

### **Distinguishing vs. eliminating possibilities, and possibilities accessible only from a first-person perspective**

Before I move on, I want to raise some concerns regarding the proposed account of self-locating knowledge. Key to Stalnaker's theory of beliefs (including self-locating beliefs) are notions of distinguishing and eliminating possibilities. It is important not to confuse these two notions.

To distinguish between possibilities is to be able to represent them (descriptively or by means of individuals), and it does not require one to know which possibility is the actual one. When I tossed a coin, before I check the result, I distinguish between two possibilities, although I cannot tell which one corresponds to the actual world. To eliminate possibilities, on the other hand, is to know which one of them is actual. The distinction between distinguishing and eliminating possibilities carries on to possibilities accessible only from the first-person perspective.

One problem related to Stalnaker's theory of *de se* beliefs is that it commits us to the view that some possibilities are accessible only from the first-person perspective. Stalnaker's solution consists in proposing a mode of identification of the thinker in terms of his occurrent thought token. This requires that we cannot have knowledge about particular thought-token other than from a first-person perspective. The claim doesn't seem controversial, but if we assume the possibility that thoughts are token-identical with physical events, it is less obvious why we cannot in principle have singular thoughts about someone else's thought tokens. If we could, we face again the possibility of error through misidentification: My belief "the person who is having *this* thought is X" does not imply a belief „I am X". We either have to rule out that thoughts are token-identical with physical events or we need to claim that there is a class of physical events which are accessible only from the first person perspective.

Second problem is how to reconcile the two claims that Stalnaker accepts: (1) a complete objective knowledge about the world requires the capacity to eliminate all possibilities which are inconsistent with the way the world actually is (i.e. one has to know the truth value of every proposition); (2) having a complete objective knowledge doesn't require one to have the capacity to eliminate these subjective possibilities (i.e. there are some propositions whose truth value one doesn't know).

One might answer that possibilities which we can represent only from the first-person perspective differ ontologically from other possibilities in that they are not real possibilities but mere representations of real possibilities. This, however, is at odds with Stalnaker's conception of possible worlds (which are not representations but real possibilities for the world, i.e. ways the world might be). Possibilities that the thinker can distinguish only from the first-person perspective should also be understood as real possibilities.

Second response one might offer is that possibilities which one cannot represent are irrelevant to one's knowledge and hence cannot be used to characterize one's epistemic situation. This is simply not true. The thinker need not have the ability to represent nor eliminate possibilities for these possibilities to be used in characterizing his beliefs. Suppose I am not aware of the existence of Plato. I cannot distinguish between nor eliminate the possibility in which Plato wrote *The Republic* and the possibility in which he didn't. It doesn't follow that this is irrelevant to my knowledge. To the contrary, we will characterize my ignorance in terms of these possibilities.

Finally, one might point out that knowledge ascriptions are context-dependent. Not being able to eliminate some possibilities in most contexts does not preclude knowledge ascription. Thus, when we say "X has all the propositional knowledge, but doesn't know whom he is" (as is the case of Lewis' gods), we are restricting the quantifier. We are leaving aside these propositions which are accessible only from the first-person perspective. This answer allows us to reconcile the two claims at the cost of a commitment to possibilities that are accessible only from the first-person perspective. Moreover, we have to accept their existence as a primitive fact.

### **The puzzle about Mary and the self-locating analogy**

How does the self-locating analogy help to solve the puzzle about Mary? In both cases – of not knowing who we are and of not knowing what it is like to experience something – the information one lacks is essentially indexical in the sense that it distinguishes between possibilities representable only from a local perspective. When we acquire a self-locating belief, we eliminate possibilities that we could represent only from a local point of view. When we learn *this is what it is like to see red*, do we **likewise** come to **eliminate** possibilities? Stalnaker answers in

the negative: he claims that just by having a phenomenal experience one acquires a cognitive ability which enables one to represent possibilities but not to eliminate them. Stalnaker's argument supporting this claim rests on the assumption that we don't gain knowledge about its essential properties just by having a phenomenal experience. It is, however, not clear how that establishes that mere knowledge about our experience does not enable us to eliminate possibilities, since: (i) In some contexts it seems natural to say that just having the experience enables one to eliminate possibilities; (ii) Stalnaker rejects the view that the thinker must know the essential properties of an object to have singular beliefs about it. Thus, in both cases (phenomenal knowledge and knowledge about external objects) whether we ascribe knowledge i.e. the ability to eliminate possibilities depends on a context. Stalnaker's claim must be weaker: just having the experience does not automatically, and in every context, amount to having the ability to eliminate possibilities.

Stalnaker presents the following thought-experiment (Stalnaker 2008:86) to support his claim that the experience itself does not enable us to eliminate possibilities: Mary is told before her release that she will be subjected to an experiment. Depending on a result of a coin toss she will be shown a red or green star. Before the experiment takes place Mary can represent two possibilities, none of which she can eliminate. After the experiment, in which she was in fact shown a red star, she still distinguishes between two possibilities, none of which she can eliminate. Mary thinks: *I know how it is to see red or how it is to see green. I don't know, which of the two colors I saw.* What has changed about Mary's epistemic situation, according to Stalnaker, is that after the experiment she is able to represent knowledge about her own experience that she couldn't represent before. She is not, however, able to eliminate possibilities. The argument doesn't seem to be conclusive: one might argue that in some contexts it is intuitive to say that the experience does enable her to eliminate possibilities (I discuss such an example below). The conception of knowledge about our experience should make sense of such cases.

### **Lewis' theory of knowledge and the principle of phenomenal indistinguishability**

Stalnaker provides an insightful critique of Lewis' theory of knowledge. He blames the inconsistencies of Lewis' view on him conflating the

externalist and internalist perspectives. On one hand, Lewis imposes very strong epistemic constraints on knowledge of objects, which makes him deny that we can have singular beliefs about them. On the other hand, he grants experience a role that it cannot play. That is because unless we accept the controversial claim that by merely having an experience we know its essential properties, our knowledge about our experience does not satisfy Lewis' restrictive criteria.

The principle of epistemic indistinguishability is the claim that worlds which are epistemically accessible to a thinker are phenomenally indistinguishable. Stalnaker defines the notion of phenomenal indistinguishability in terms of a cognitive capacity: two mental states are phenomenally indistinguishable iff the subject can switch from one to the other without noticing any difference. Stalnaker claims that the thought experiment with Mary and the two stars shows that the principle is false. After being presented with the red star, Mary still doesn't know which one is the actual world: the one in which she was shown a red star, or the one in which she was shown a green star. The two worlds should be phenomenally indistinguishable for her, while in fact they are not. If they were, we would have to accept that there is a counterfactual world in which she saw a green star, which is phenomenally indistinguishable from the actual one in which she saw a red star. The principle of epistemic indistinguishability commits us – Stalnaker concludes – to the existence of phenomenal information (Stalnaker 2008: 90-91).

We may, however, disagree with Stalnaker in that Mary doesn't know whether she was shown a red or green star. She knows which star she was shown, she only doesn't know the name of its color. If she was first shown a red star, and a moment later a green, yellow, and a blue one, and if she was asked which of the stars she saw first, she would be able to eliminate the possibilities. Stalnaker does indeed discuss a similar case: he claims that even when Mary names all the colors, but she cannot relate these names with the names she was using while locked in the black and white room, it is a matter of context whether we would ascribe to her knowledge that the object is red. In some situations of this type we tend to say that Mary knows that this tomato is red, and in another we don't (when we want to emphasize the fact that she is unable to connect her old concepts to her new experiences). Stalnaker, however, thinks that the fact that in some contexts we would be reluctant to ascribe

knowledge is enough to undermine the principle of phenomenal indistinguishability.

Stalnaker presents us with the following choice: either we stick to the principle of phenomenal indistinguishability at the cost of accepting that phenomenal information exists or we reject the principle and accept that knowledge about our experience is not epistemically privileged and does not play the role it was granted by empiricist epistemology. Stalnaker argues for the latter option: the knowledge about our experience is as indirect as knowledge about external objects. What he attempts to do, however, is to elucidate the notion of direct knowledge by explaining the intuitions that motivate it.

Lewis imposes a very strong epistemic constraint on the knowledge of objects: singular thoughts about objects require knowledge of their essential properties. Since we don't know essential properties of objects, Lewis claims that we cannot have singular thoughts about them. He uses Saul Kripke's puzzle about Pierre to justify this claim (Lewis, 1981). Pierre, as we remember, thinks about London (when he is still in France and calls it "Londres") that it is pretty and (when moving to London and using its English name) that it is not pretty. If we accept the theory of direct reference and the disquotational principle we are forced to ascribe to him contradictory beliefs. That violates our intuition that Pierre is rational. Lewis argues that the puzzle lends support to the internalist theory of beliefs. He argues that we might conceive of a situation in which the French name "Londres" designates a different city than London, Bristol for instance. Such a world is for Pierre indistinguishable from the actual one. Since Pierre doesn't know London's essential properties, there are such possible worlds epistemically accessible to him in which the name "Londres" refers to a different city than London. Whenever we have beliefs about objects whose essential properties we don't know, there are epistemically accessible worlds in which some other object plays the same role as the given object plays in the actual world.

An anti-individualist (an externalist) would object that the belief Pierre would have in that counterfactual world differs from the one he has in the actual world. Lewis rejects this counterargument claiming that we need a narrow notion of content in order to explain how we can have access to our own beliefs and to avoid the conclusion that Pierre is irrational.

Lewis claims that we cannot have a singular thought about an object without knowing its essential properties, and the same goes for our knowledge about phenomenal experience: we cannot have singular beliefs about our experience without knowing its essential nature. Lewis thus rejects the controversial claim that we know the essential properties of a phenomenal experience merely by having this experience. Accepting this controversial claim amounts to saying that by merely having an experience we would eliminate all possibilities in which my phenomenal experience has different physical nature. As Stalnaker points out, a materialist cannot accept this claim, as he holds that experiences are identical to physical states and that it is physical properties that are essential to them. We don't know the physical nature of our experiences merely by having them. Having a phenomenal experience of a given type, for instance a headache, is not sufficient. I do not know whether my experience is a complex physical state of type A or B.

Stalnaker points out that the privileged role that experience plays in Lewis' theory of knowledge is at odds with his rejection of the claim that just by having an experience one knows its essential properties (Stalnaker 2008: 99). According to Lewis' theory, knowledge is represented by possibilities which are not eliminated by experience (Lewis, 1996). A possible world  $w$  is not eliminated by experience iff the subject's perceptual experience and memory in  $w$  are the same as they are in the actual world. Lewis gives a contextualist response to a skeptic's concern "how then can we have knowledge which goes beyond our experience?": depending on the context, we are allowed to ignore some possibilities which our experience does not eliminate. Let's illustrate this idea with an example. John knows that it rains iff his experience eliminates every possibility in which it does not rain. We can, however, imagine that what John takes to be rain is an effect produced by a film crew. Although John's experience does not eliminate this possibility, in a normal context (when it actually rains) this does not preclude us from ascribing knowledge to him.

However, what is of key importance to Lewis' theory is the assumption that we will know that the possibilities eliminated by experience, in any context, are inconsistent with our knowledge. This – Stalnaker argues – implies that by having an experience we gain knowledge about its essential properties and hence implies the above

mentioned thesis (that having an experience is knowing its essential properties), which Lewis in fact rejects. Accepting this controversial thesis, as Stalnaker reminds us, implies the phenomenal indistinguishability principle together with its correlate, the existence of phenomenal information.

Stalnaker blames these inconsistencies in Lewis' theory on him conflating the externalist and the internalist perspectives. The theory is externalist in that it describes the experience from an external perspective: as a set of possibilities in which the subject's experiences are identical. It is, however, internalist in the privileged role of experience (Stalnaker 2008: 101). Lewis' theory identifies knowledge with possibilities not eliminated by experience. As Stalnaker points out, according to Lewis' theory, it is the mere occurrence of an experience that eliminates possibilities and not its propositional content. Worlds which are not eliminated by experience are worlds in which the subject has the same experiences as in the actual world. What the theory assumes is that two identical experiences have identical causes, hence there is a one-to-one relationship between phenomenal properties of our experience and the physical features of the world which cause this experience.

Stalnaker is critical of these ideas. He thinks that Lewis misconceives the role of experience in eliminating possibilities by not recognizing that whether we'll say that a certain experience eliminates possibilities or not depends on a context. Let's think once again about Stalnaker's thought experiment with Mary seeing a red or a green star. After being shown a red star, Mary is still not able to eliminate one of the two possibilities. But whether we'll judge her to know which star she has seen (and ascribe her the ability to eliminate possibilities) depends on our, the theorist's aims. Imagine two scenarios: in the first scenario we tell Mary that first she will be shown a red or a green star and later she will be presented with two buttons, a red and green one. If she presses the red button, the world will be annihilated. If she presses the green one, nothing will happen. In the second scenario, we tell her that later she will be presented with two buttons, a red and green one. If she presses the button of the color that she was exposed to during the experiment, the world will be annihilated. In the first case, we may say that the mere experience did not enable Mary to eliminate possibilities. In the second case, we will say that having the experience she has eliminated a

possibility. Thus, knowledge we ascribe on the basis of the experience depends on the context of ascription.

Lewis' contextualism is – according to Stalnaker – not thorough enough. Although Lewis thinks that in our everyday practice of belief ascription we are entitled to eliminate certain possibilities which seem irrelevant, we might still speak about knowledge in the absolute sense. Stalnaker disagrees with Lewis on this point. He also argues that it is not the mere occurrence of an experience which eliminates possibilities but their propositional content. As he points out, there is no direct connection between having an experience and the propositional content of our beliefs.

### **The principle of epistemic transparency and anti-individualism**

Many theorists think, like Bertrand Russell, that having a singular belief about an object requires that the thinker is in a special epistemic relation with the object. Without our being acquainted with the object, the object cannot be the content of our beliefs. There is no single interpretation of what the relation of acquaintance amounts to. One possible interpretation of this notion was proposed by Lewis: in order to be acquainted with an object one has to know its essence (Lewis, 1981). Since we don't know the essential properties of objects and persons, a proponent of this view has to deny that we have singular beliefs about them. The problem generalizes our knowledge of properties and relations. Stalnaker demonstrates that, contrary to the traditional view, we are not in such a privileged relation with our experience.

The key idea to Stalnaker's theory of belief ascriptions is his "deep contextualism". It is the view that there is no one correct context-independent characterization of the thinker's beliefs. It is not the case that all context-dependent characterizations should be regarded as a mere approximation which could after all be substituted with correct context-independent characterizations (which we don't do for practical reasons). Stalnaker claims that contents are essentially ascribed and not inherent. On this theory, having singular thoughts is not a matter of a thinker's acquaintance with an object or his having a particularly rich conception of the object. Whether we characterize one's beliefs in terms of the object singularly or descriptively depends on the context of the ascription. Stalnaker demonstrates that we might have a very detailed conception of an object and yet not be aware that this conception refers

to one and the same object. In order to characterize such a belief, we'll need to do it descriptively. On the other hand, in a different context, one might know very little about an object beyond some contingent fact, and still we might be able to ascribe to him a singular belief about this object. It is the context of ascription that is critical to what we'll judge as a correct characterization of someone's beliefs.

Deep contextualism also provides a way to accommodate a popular intuition that for our thoughts to explain the agents' actions they must be epistemically accessible to them. Many authors have claimed (for example Paul Boghossian) that the intuition cannot be reconciled with anti-individualism, which holds that it is facts about the environment which determine the contents of our thoughts. Stalnaker thinks that an anti-individualist can make sense of the former intuition if he accepts the view that contents are externally ascribed by the theorist.

The principle of transparency, which expresses the above mentioned intuition, says that for two thoughts of a subject to have the same content the thinker must know *a priori* that it is so (likewise if his two thoughts have different contents, the thinker must know that it is so) (Boghossian, 1994). The conflict between anti-individualism and the principle can be seen in the case of contradictory beliefs: situations in which we hold contradictory beliefs about an object, not realizing that our thoughts refer to one and the same object or situations in which we have beliefs about an object, not being aware that our beliefs refer to two different things. In these situations we cannot ascribe beliefs in a way prescribed by the anti-individualist theory or we would have to conclude that the subject is irrational.

One type of arguments against anti-individualism makes use of the so called 'slow-switching scenario' (Boghossian, 1994). It is a thought experiment which involves a thinker being transported from one context to another (from Earth to Twin Earth) in such a way that he is not aware of the change and the two contexts are indistinguishable for him. The thinker's beliefs on Earth are about water. When he thinks, shortly after being transported to Twin Earth, "There is water in this lake", according to an anti-individualist, he has a false belief, which concerns water and not a true belief concerning XYZ. At the same time, we have a strong intuition that after years spent on Twin Earth, when having this type of thought, he no longer thinks about water, but about XYZ. He can also – to the detriment of our theories – compare his earlier beliefs with his recent

ones: while living for many years on Twin Earth he might recall: *once* (thinking of a specific episode from his Earthly life) *water tasted much better than nowadays*.

Thought experiments of this kind demonstrate that our intuitions concerning the contents ascribed change with the situation. The challenge they are meant to pose for an anti-individualist is that he should account for the mysterious change that the subject undergoes which makes his thoughts change reference. Stalnaker rejects this way of putting the problem as deeply misleading (Stalnaker 2008: 121). Instead of situating the change in the subject (in his head), we ought to explain our intuition that the reference of his thoughts changes in terms of the context in which we make these belief ascriptions. Generally, we ascribe beliefs to explain and predict the agents' behavior. It follows that we must ascribe beliefs in a way which is consistent with the assumption that the agent is a rational being. The principle of transparency reflects a different requirement: not on the subject, but on the theorist: the ascriptions that the latter makes cannot violate the subject's rationality assumption.

Stalnaker convincingly shows that the principle of epistemic transparency can be reconciled with externalism, understood however, I would argue, slightly differently than the early formulations of the view (Burge, 1979) suggest. Burge claimed about two thinkers, who (i) live in two linguistic communities that differ only in their use of the term "arthritis" and (ii) nevertheless associate the same set of descriptions with the term arthritis, that they do not have any common beliefs about the disease. Externalism, as it is interpreted by Stalnaker, does not exclude the possibility that in some particular context it might be perfectly fine to characterize their beliefs with the same sets of possibilities. Stalnaker's externalism claims that (1) the content of our mental states depends on the facts concerning our environment in the sense that we should look for an explanation of why we have beliefs of that content in our causal relations with the environment. He also states that (2) contents are essentially ascribed and not inherent and that these ascriptions are made in a particular context which includes the cognitive aims of the theorist who does the ascription. The theory is also externalist in the sense that when making ascriptions we use the resources which are available to us: that is objects, properties, and relations which are there in the actual world.

Stalnaker's book makes a strong case for externalism understood as a methodology rather than a metaphysical view. At the same time, he acknowledges important internalist intuitions (e.g. that we perceive the world from a certain point of view,) and shows that, by reversing the order of explanation, we can do justice to them on externalist grounds. Stalnaker shows that the possible worlds representation of content enables us to represent the subjective point of view of the thinker as well as the relation between his perspective and the way the world is in itself. The book is both very rewarding and very demanding. For, although Stalnaker avoids technical details, he connects variety of philosophical issues, often shifting the grounds of the discussion.

### **Acknowledgments**

I am grateful to Tadeusz Ciecierski for his comments on the earlier draft of this paper.

## REFERENCES

- Burge, T. (1979), "Individualism and the Mental", *Midwest Studies in Philosophy* 4: 73-121.
- Elga, A. (2000), "Self-Locating Belief and the Sleeping Beauty Problem", *Analysis* 60:143-147.
- Lewis, D. (1979), "Attitudes *de dicto* and *de se*", *Philosophical Review* 88: 513-543.
- Lewis, D. (1981), "What Puzzling Pierre Does not Believe", *Australasian Journal of Philosophy* 59: 283-289.
- Lewis, D. (1988), "What Experience Teaches", in Lewis, D. (1999) *Papers in Metaphysics and Epistemology*, Cambridge: CUP.
- Lewis, D. (1996), "Elusive Knowledge", *Australasian Journal of Philosophy* 74: 594-567.
- Nida-Rümelin, M. (1995), "What Mary Couldn't Know: Belief and Phenomenal States", in Metzinger, T. (ed.), *Conscious Experience*, Exeter: Imprint Academic: 219-241.
- Perry, J. (2001) *Knowledge, Possibility and Consciousness*, Cambridge, Mass.: MIT Press.
- Stalnaker, R. (2008) *Our Knowledge of the External World*, Oxford: OUP.

## ABSTRACT

### KNOWLEDGE ABOUT OUR EXPERIENCE AND DISTINGUISHING BETWEEN POSSIBILITIES

In my article I reconstruct the main threads of Robert Stalnaker's book *Our Knowledge of the Internal World*, which focuses on the problem of our epistemic relation to our experience and the relation between experience and knowledge. First, the book proposes an interesting view of externalism, which combines classical externalist claims with a contextualist approach to content ascriptions. The approach accommodates some important internalist intuitions by showing how content ascriptions can be sensitive to the perspective from which a subject perceives the world. Second, Stalnaker proposes a theory of self-locating and phenomenal knowledge, which should be understood in terms of differentiating between real possibilities. The puzzling upshot of this elegant solution is that it commits one to the existence of possibilities accessible only from the first-person perspective. Finally, Stalnaker presents an argument which shows that our knowledge about our phenomenal experience is no more direct than the knowledge about external objects. Stalnaker's claim that by merely having an experience we don't learn any new information seems, however, too strict in light of his contextualist approach to content ascriptions.

**KEYWORDS:** Robert Stalnaker; externalism; contextualism; phenomenal experience; self-locating beliefs