

In: Allan, K., Capone, A. and Kecskes, I. (eds.): *Pragmemes and Theories of Language Use*. Cham, Heidelberg, New York, Dordrecht and London: Springer, 2016, 643-677.

Poor vs. good thought experiments in pragmatics: A case study¹

András Kertész

MTA-DE Research Group for Theoretical Linguistics
Hungarian Academy of Sciences
and
University of Debrecen
Institute of German Studies
Pf. 47
H-4010 Debrecen, Hungary
andraskertesz@yahoo.de

Abstract

Thought experiments have played a decisive role in the development of pragmatics, and yet the problem of how to distinguish between good and poor thought experiments has not been tackled systematically. The present paper is devoted to this problem. After presenting a brief overview of the philosophical literature on the distinction between good and poor thought experiments, as a point of departure the author defines the Received View. This says that a thought experiment is poor if (a) it leads to contradictory conclusions, and/or (b) it is fallacious (with special respect to circularity), and/or (c) it cannot be related to real experiments/empirical theories, and/or (d) it misses relevant experiential content. The present paper puts forward another hypothesis which is the result of applying a novel metatheoretical model of plausible argumentation to the problem mentioned. According to the new hypothesis, a thought experiment in pragmatics is poor if it does not trigger a process of plausible argumentation that is cyclic, prismatic and based on the retrospective re-evaluation of information. The tenability of this hypothesis is exemplified by a case study on one of Searle's famous thought experiments.

Keywords: thought experiment, real experiment, plausible argumentation, meaning

1 Introduction

One important aspect of Professor Mey's seminal achievements is his careful treatment of different data types and data sources. On the one hand, he investigated authentic empirical data rooted in written sources such as newspapers and literary texts as well as oral conversations. On the other hand, he also made use of the very data type that was the major

¹ I am grateful to Katinka Halász for her insightful comments on earlier drafts. I alone am responsible for the shortcomings of the present paper.

object of investigation in the philosophically oriented classic works of those who shaped the birth of "the 'young' science of pragmatics" (Mey 1993: 3): thought experiments. That professor Mey is well aware both of the use of thought experiments in pragmatics and their interplay with other data types, can be nicely illustrated by the way he explains the notion of 'pragmatics' on the first pages of his widely used textbook (Mey 1993). On p. 4 Professor Mey begins his answer to the question of 'What is pragmatics all about' with analysing a passage from the Chicago weekly *Reader*, and a few pages below he exemplifies one of his claims by a quotation from David Lodge's *Paradise News*. However, on pages 8-9 the reader will find another kind of data, namely a typical thought experiment: a witty imaginary dialogue between two linguists, called Jacob and Mark. How Professor May, despite the fact that the main data sources of his research were authentic texts, appreciates the power of thought experiments, is spectacularly witnessed, for example, in his 'Pragmatic Comedy in Five Acts' which starts with a scenery the protagonists of which are Fridolin, the poet, and his partner, a peasant (Mey 1987). These examples show that Professor Mey's work is an excellent motivation for devoting a couple of thoughts to the nature of pragmatic thought experiments.

Although thought experiments have played a decisive role in the development of pragmatics and have prepared many of its central findings, to our knowledge their nature has not as yet been investigated systematically. They give rise to numerous questions that pertain to the foundations of pragmatic theorizing, because they have been frequently used as tools of pragmatic research and have triggered many seminal findings. One of the questions that has not been raised but that seems to be at the heart of the foundations of pragmatics concerns the distinction between good and poor thought experiments. Thus, the present paper will be devoted to the following problem:

(P) How to distinguish good from poor thought experiments in pragmatics?

(P) is motivated not only by the impact of thought experiments on pragmatics, but by general discussions in the philosophy of science as well. Namely, during the past decades, the nature of thought experiments has become one of the central topics in the philosophy of science. A vast literature on the subject has been produced and there have been challenging discussions on this topic.²

In order to find a suitable point of departure for tackling (P), let us start with a brief overview of the answers that have been given to the question of how to distinguish between good and poor thought experiments in natural science and in philosophy.³ Fortunately, although the views on the nature of thought experiments diverge, the answers to the latter question seem to converge. Therefore, we will try to identify the main common components of these answers and, on this basis, to define the Received View.

To begin with, we refer to the classic work of Sorensen (1992). In thinking over the difference between good and bad thought experiments, Sorensen advocates a broad and sophisticated view. First, he claims:

² For overviews of the philosophical literature on thought experiments see Brown (1991), Brown and Fehige (2014), Sorensen (1992), Kühne (2005), Buzzoni (2008), Cohnitz (2006a), Gendler (2000), Moue et al. (2006).

³ We do not attempt to define the notion of 'thought experiment' here. On the one hand, Kertész (2015) and Kertész and Kiefer (2013) highlight some of their central features within Kertész and Rákosi's (2012) framework. On the other hand, as Peijnenburg and Atkinson (2003: 306) maintain, "we do not feel the need to state exactly what thought experiments are; after all one can distinguish good from bad theories, or thoughts, or experiments without being able to define what exactly theories, thoughts or experiments are."

"Since thought experiments also have a function, we can say that a good thought experiment *efficiently raises or answers its questions*. Thought experiments are *intellectual instruments* like arguments, demonstrations, explanations, and research programs. They are supposed to make you an authority by having you merely reflect on the experimental plan. Hence, thought experiments will be prized to the extent that they enlighten in this minimalist fashion." (Sorensen 1992: 252; emphasis added).

The quotation highlights the *heuristic function* of thought experiments. Accordingly, they have to be evaluated relative to the effectiveness of their problem solving capacity. Second, Sorensen specifies the above claim in maintaining that the reason why thought experiments may fail is that the thought experimenters commit *fallacies*.⁴ After delimiting fallacies from what he calls 'myths' and 'abuses', he adds that

"[a]ny fallacy can be committed in a thought experiment – or ordinary experiment for that matter. Hence, there are no fallacies *unique* to thought experiments just as there are no diseases unique to miners." (Sorensen 1992: 256; emphasis as in the original)

While Sorensen discusses a series of fallacious and antifallacious thought experiments in the natural sciences and in philosophy, the current literature focuses on special aspects that the authors deem especially relevant. We illustrate this by three examples taken from present day approaches.

As the first example, we cite Peijnenburg and Atkinson (2003).⁵ The authors draw a dividing line between scientific and philosophical thought experiments and claim that the indications of poor thought experiments are more devastating in the latter case than in the former. Both in the natural sciences and in philosophy, one characteristic of a poor thought experiment can be *inconsistency*: the thought experiment leads to contradictory conclusions in the sense that different people can draw incompatible conclusions from the same thought experiment. The second indication is one of the classical fallacies, namely, begging the question. Thereby, the argumentation that the thought experiment hinges on is *circular* in that the conclusion just reiterates what has already been assumed at the outset. However, scientific thought experiments may avoid the disaster that such shortcomings anticipate insofar as they can be related to empirical theories and/or real experiments. As opposed to this, philosophical thought experiments are fully exposed to the consequences of inconsistency or circular argumentation, because the findings they suggest cannot be controlled by real experiments and/or empirical theories.

With respect to our problem (P), one important implication of Peijnenburg and Atkinson's approach is that although they only speak of the two indications mentioned, there seems to be a third indication of poor thought experiments as well. Namely, it follows from their reasoning that a thought experiment may be poor if it cannot be related to certain kinds of empirical theories and/or real experiments – and this should apply both to scientific and philosophical thought experiments. Another implication relevant for (P) is that if we accept the authors' claim that the consequences of being poor are less catastrophic for scientific thought experiments than philosophical ones, then – because of their philosophical origin – the classic thought experiments in pragmatics appear to be especially challenging. In particular, on the one hand, the classic thought experiments that fertilized the emergence of pragmatics were conducted by philosophers such as Austin, Grice, Searle, Wittgenstein and others, and are thus rooted in the philosophy of language. On the other hand, they also

⁴ "A fallacy is a bad inference rule that looks like a good one. An antifallacy is a good inference rule that looks like a bad one." (Sorensen 1992: 274)

⁵ See also the discussion that Peijnenburg and Atkinson's (2003) paper raised (Cohnitz 2006b, Peijnenburg and Atkinson 2007).

triggered the abrupt development of pragmatics as a field which is on the way to becoming an experimental discipline; in fact, current experimental pragmatics – not exclusively, but at least to a considerable extent – focuses on the empirical testing of ideas first put forward as findings supported by the classic thought experiments of philosophers. Then, presupposing the difference that Peijnenburg and Atkinson assume between scientific and philosophical thought experiments, the question as to the consequences of a pragmatic thought experiment's being evaluated as poor is of particular importance.

Rachel Cooper's (2005) approach is our next example. She claims that thought experiments may fail in two ways. First, it may be the case that the thought experimenter assumes that in a particular situation there is an inconsistency where there is not, and this misrepresentation motivates the thought experiment. The content of such inconsistency is different from what Peijnenburg and Atkinson discuss: while the latter pertains to the consequences yielded by the thought experiment, the former stems from the motivation of the thought experiment. The second way a thought experiment may fail is closely connected to the fact that thought experiments are assumed to include explicit or tacit *experiential information* of the world. Then, a thought experiment may fail if it does not contain relevant experiential information. Cooper generalizes these two cases in concluding that

"[i]n general, we can say that a thought experiment is more likely to succeed if the thought experimenter is knowledgeable about the relevant aspects of the actual world. Only if she possesses either explicit or implicit knowledge of the behaviour of real phenomena can the thought experimenter predict how hypothetical events would unfold." (Cooper 2005: 343)

The third example we mention is Camilleri (2014) which does not focus on the question concerning the failure of thought experiments explicitly, but which still implies a clear answer to it. In particular, Camilleri emphasizes that "the ability to execute a TE [thought experiment] well depends on the *ability to recognize* that some piece of previously acquired knowledge is *relevant* to the situation at hand" (Camilleri 2014: 1711; emphasis as in the original). Consequently, a thought experiment may be poor if this kind of relevant information about the world is missing. Thereby, in both Cooper's and Camilleri's approaches a wide range of possible kinds of experiential information are allowed: what is relevant for the thought experiment may be explicit knowledge of laws or tacit knowledge of them, or simply personal experiences with objects or situations in the world.

Having exemplified typical views put forward in the literature, we may now summarize what we will call *the Received View* of the failure of thought experiments:

- (RV) A thought experiment is poor if it
- (a) leads to contradictory conclusions, and/or
 - (b) is fallacious (with special respect to circularity), and/or
 - (c) cannot be related to real experiments/empirical theories, and/or
 - (d) misses relevant experiential content.

In the present paper we will put forward a different hypothesis as our own solution to (P):⁶

⁶ Although, for example, Peijnenburg and Atkinson (2003: 305) assume that there are "criteria which distinguish good thought experiments from bad ones", they themselves call inconsistency and begging the question not 'criteria' but merely 'indications'. We will evaluate (H) in a similarly cautious way: in order to decide whether (H) provides a necessary and/or sufficient criterion, a great number of thought experiments ought to be analysed, which would go far beyond the present paper.

- (H) In pragmatics a thought experiment is poor if it does not trigger a process of plausible argumentation that is cyclic, prismatic and based on the retrospective re-evaluation of information.

Our line of reasoning will be as follows:

In Section 2, we will introduce a framework of plausible argumentation called the p-model (Kertész and Rákosi 2012) that yields (H). In Section 3, we will illustrate the workability of the p-model and the way it implies (H) by applying it to a case study on Searle's classic thought experiment centred on Goethe's question *Kennst du das Land, wo die Zitronen blühen?*⁷ Thereby, with the help of the p-model, we will examine (RV) (a), (c) and (d); that is, we will check to what extent this thought experiment can be evaluated as being poor in the sense of (RV). We will not discuss (RV)(b), because, although in a partly different context, the circularity of thought experiments in pragmatics – as a particularly important kind of fallacy – as well as its consequences, have already been exemplified by 'Grice's circle' in Kertész and Kiefer (2013).⁸ Finally, Section 4 will summarize our findings.

2 Basic notions of the p-model⁹

The central hypothesis of Kertész and Rákosi's (2012) p-model says that linguistic theories are processes of plausible argumentation.¹⁰ The main motivation for the p-model is the insight that the way language structure and language use are investigated in different branches of linguistics can be understood only if the *essential uncertainty* of linguistic theorizing is accounted for. It is this uncertainty that shapes pragmatic research as well. Accordingly, the p-model is intended to capture the following key properties of pragmatic theorizing: the *uncertainty* of the data; drawing *plausible inferences* from uncertain premises instead of deductive and/or demonstrative ones; the *cyclic and prismatic nature* of pragmatic theorizing based on uncertain data; the way the *inconsistency* between the hypotheses and the data or between rival hypotheses is treated; and the *pluralism* of pragmatic theorizing. Due to lack of space, here we will not give a comprehensive overview; rather, we will introduce in an informal manner only those notions that are needed for the solution of the main problem of the present paper.

Plausible statement. The p-model decomposes statements into two components. One is its informational content, and the other is the plausibility value that indicates to what extent the information content of the statement is supported by a given *source*, that is, to what extent the source on the basis of which the statement is accepted counts as *reliable*. One is ready to accept the statement on the basis of this source.¹¹

Sources. There are two kinds of sources of the plausibility of the statements at issue. Physical or intellectual sources (sensual experiences, observations, papers, books, experiments, intuition, corpora etc.) are called *direct sources*. For example:

⁷ See the Appendix of the present paper.

⁸ In addition, the way the p-model treats fallacies was shown in Kertész and Rákosi (2009).

⁹ For concise overviews, see Kertész (2015) and Kertész and Rákosi (2014a), (2016). For the detailed discussion and the precise definition of these notions, see Kertész and Rákosi (2012). For seminal approaches to plausible inferences, which also fertilized the p-model, see Rescher (1976), (1987), Polya (1948), (1954) and Walton et al. (2008).

¹⁰ Several case studies have been published in which the p-model is applied to different fields of linguistics. Cognitive linguistics: Kertész and Rákosi (2009), Kertész (2015); Pragmatics: Kertész and Kiefer (2013), Kertész and Rákosi (2014b), Kertész and Rákosi (2016); Syntax: Kertész and Rákosi (2013). See also the case studies in Kertész and Rákosi eds. (2014).

¹¹ Please notice that it is not the statement but its plausibility value that has a source in this sense.

- (1) By asking the question 'Can you check my paper?', Andrew implicated that he would like Katie to check his paper.

The claim in (1) may appear in a scholarly paper on pragmatics. It may be deemed plausible on the basis of the scholar's pragmatic intuition as a direct source. If the plausibility value of a statement is supported by other statements the plausibility value of which has already been known, then we speak of an *indirect source*. As a rule, indirect sources are *inferences*.

Plausibility values. We will indicate the plausibility value of statements by putting them within '|' and referring to the source at issue in the subscript. For example, '| p | S ' is to be read as 'the plausibility value of the statement p on the basis of the source S '. The plausibility value of a statement can be characterized in general as follows:

- (2) (a) $|p|_S = 1$, if p is true with certainty on the basis of S ;
 (b) $0 < |p|_S < 1$, if p is plausible on the basis of S ;
 (c) $0 < |\sim p|_S < 1$, if p is implausible on the basis of S ;
 (d) $|p|_S = 0$, if p is of neutral plausibility on the basis of S , i.e., if it is neither plausible nor implausible on the basis of this source.

For example (1), supplemented by the notation thus introduced, will look like this:

- (3) $0 < |$ By asking the question 'Can you check my paper?', Andrew implicated that he would like Katie to check his paper. $|_S < 1$

Here S may stand for the pragmatic intuition of the researcher as a direct source. Plausibility values are not absolute but source-dependent and merely serve the comparison of statements. It is often the case that a particular statement is more plausible than another on the basis of a certain source but less plausible with respect to another source.

Plausible inference. Inferences whose premises do not prove the truth of the conclusion but partially support its plausibility are called plausible inferences. The premises of a plausible inference are plausible statements that are related to the conclusion by some semantic relation such as causality, analogy, whole-part relation, part-whole relation etc. Plausible inferences need not follow the patterns of deductively valid inferences, but they may be *enthymematic*.¹² In enthymematic inferences there are missing *latent background assumptions* whose addition would make the structure of the inference deductively valid. For example:

- (4) Premises:
 (a) $0 < |$ Andrew's body is big. $|_{S1} < 1$
 (b) $0 < |$ Andrew's head is part of his body. $|_{S1} < 1$
 Conclusion:
 (c) $0 < |$ Andrew's head is big. $|_{I(4)} < 1$
- (5) Premises:
 (a) $0 < |$ Andrew's body is big. $|_{S1} < 1$
 (b) $0 < |$ Andrew's head is part of his body. $|_{S1} < 1$
 (c) $[0 < |$ If the whole has property P , then the part also has property P . $|_{S2} < 1]$

¹² According to Walton et al. (2008: 189), "[o]ne problem with enthymemes is that reasonable people can have differences of opinion on what the implicit assumptions are supposed to be. Filling in the missing parts of an enthymeme may depend on interpreting the natural language text in which the argument was put forward [...]."

Conclusion:

(c) $0 < | \text{Andrews's head is big} |_{I(5)} < 1$

In (4) and 5, $I(4)$ and $I(5)$ refer to the indirect source of the conclusion i.e. the inference in (4) and (5), respectively. (5) is the supplementation of (4) by the latent background assumption in (5)(c).¹³ It may be plausible but need not be true.

The p-context. The p-context is that subset of information at one's disposal that is related to the plausibility value of relevant statements. It comprises those reliable sources on the basis of which the plausibility value of the statements can be evaluated; the plausible statements; and the accepted methodological principles. A p-context may be *informationally overdetermined* which means that it includes too much information in the sense that there is at least one statement that is plausible on the basis of a certain source but implausible on the basis of another within the given p-context. We call such cases *p-inconsistency*. The p-context may be *informationally underdetermined*, if none of the known sources is capable of assigning a plausibility value to a given statement. A p-context may be both under- and overdetermined with respect to different statements.

P-problems, their solution and their resolution. We will say that a p-context is *p-problematic* if it is informationally over- and/or underdetermined, and instances of informational over- or underdetermination will be called *p-problems*. We will call a p-context in which the statement in question is either plausible or implausible with respect to all sources supporting it, *the solution of a p-problem*. Nevertheless, it may be the case that a p-problem has several solutions rooted in different sources. The most acceptable one of these solutions is *the resolution of the p-problem*.

Plausible argumentation. Plausible argumentation is a process consisting of sequences of plausible inferences. It is a *heuristic tool* that aims at diminishing the informational over- and/or underdetermination of the p-context, i.e. the resolution or at least one possible solution of the given p-problem. Normally, it is not linear, but includes *cyclic* returns. The cycles are also *prismatic* insofar as they continuously change the perspective from which the information in the p-context is evaluated. Thereby, previous decisions are *retrospectively re-evaluated*.¹⁴

The phases of the plausible argumentation process. The plausible argumentation process begins with what we call the *starting p-context* that includes, among others, the p-problem the argumentation process is devoted to, and terminates with *the final p-context* resulting in the (at least provisional) solution or resolution of the p-problem. Between these two phases there may be a series of *argumentation cycles*. An argumentation cycle is a particular phase of the argumentation process that leads to the (provisional) solution or resolution of the p-problem at issue and it may include several subcycles. In the prototypical case (from which particular cases of plausible argumentation may differ), a cycle consists of three main stages. The first is the construction of a *new p-context version* that may include the original p-problem or a particular version of it as well as the plausible statements, the latent background assumptions, and the methodological principles needed for the solution or the resolution of the p-problem. The next stage is *the elaboration of the p-context* in the course of which new sources, new statements, or new methods are introduced, or already accepted ones modified or deleted. The third stage is the *comparison* of the p-context versions with their rivals, whereby it has to be decided which of them represents the most satisfactory solution. If no decision can be made or the decision is unsatisfactory, then the argumentation process may return to the starting p-context and a new argumentation cycle begins, and so on.

Data. The p-model defines linguistic data as plausible statements with a positive

¹³ Latent background assumptions are set within square brackets.

¹⁴ See also Rescher (1987) for these features.

plausibility value that is rooted in a direct source. Thus, it is not the linguistic phenomena themselves – sentences, utterances, texts, moves in a dialogue – that count as data, but rather, the plausible statements about them. For example:

- (6) (a) *Katie checked Andrew's paper.*
- (b) $0 < | \text{The sentence } \textit{Katie checked Andrew's paper} \text{ is grammatical.} |_S < 1$
- (d) $0 < | \text{The sentence } \textit{Katie Andrew's paper checked} \text{ is ungrammatical.} |_S < 1$

(6)(a) is not a datum in this sense. (6)(b) and (c) are data stemming from the direct source *S* which in this case may be the linguistic intuition of a native speaker.

Having introduced the concepts we will need, we now turn to a case study exemplifying (H).

3 On (RV)(a), (c) and (d): A case study

3.0 Introductory remarks

Before starting the case study, we note the following:

First, Searle's thought experiment mentioned in Section 1, provides an instructive illustration of our solution to the problem (P), because at first sight it seems to be a clear case of a poor thought experiment in the sense of (RV): it gives rise to contradictory conclusions drawn by different people; when it appeared, it was a contribution to the philosophy of language and had nothing to do with real experiments and/or empirical theories; and it includes an imaginary story with a counterfactual conclusion that tells us what the case would be if in World War II, Searle, as an American, were captured by Italian soldiers and quoted Goethe's famous poem in German.

Second, we distinguish between the notion of 'thought experiment' and that of 'thought experimental report' – in analogy to the distinction between 'real experiment' and the corresponding 'experimental report'. A thought experiment is a non-public mental process, while a thought experimental report is a text making certain elements of the non-public mental process public. A real experimental report is a public text as well, while the experiment it refers to is non-public in that it is restricted to a small community of experimenters involving the complicated interplay of mental, communicative, social and technical processes. It is the thought experimental report quoted in the Appendix that is the subject of our analysis and not the thought experiment itself. Similarly, we will analyse reports on experiments as well, but not the experiments directly. However, although they do not have the same structure in every respect, thought experimental reports and experimental reports include pieces of information the origin of which is the original thought experiment or real experiment, respectively. In the thought experimental report or experimental report these pieces of information are systematized into a public argumentation process. Therefore, from the argumentation structure of the thought experimental/experimental report one may infer properties of the thought experiment or real experiment at issue. Throughout the present paper, we will use the terms 'thought experiment', 'real experiment' (or its usual short form: 'experiment'), 'thought experimental report' and 'experimental report' consistently, but, in order to keep our train of thought within reasonable limits, we will not go into the explication of the highly complex relations between them.¹⁵

Third, what we aim at in the case study is a *partial metatheoretical reconstruction* of the way Searle uses his thought experimental report in order to argue against Grice's theory

¹⁵ See Kertész and Rákosi (2014b) for the discussion of these relations.

and for his own. It is *not* our aim to take sides in the discussion between them and *we will not* argue for Searle's standpoint and against Grice's view of meaning.

Fourth, we do not analyse Searle's thought experiment for its own sake; the focus of the paper is not Searle's thought experiment, but (H). The thought experiment merely serves to exemplify that instead of (RV)(a), (c) and (d), it is the hypothesis (H) that is more favourable. We could have chosen a series of other thought experiments as possible illustrations of the tenability of (H).

Finally, the present paper continues earlier research that aimed at the clarification of different aspects of the structure and function of thought experiments in certain fields of linguistics. For lack of space and to avoid the repetition of our findings already published, we do not intend to give a comprehensive account of thought experiments in pragmatics; we restrict our attention to (P) and its possible solution (H).

3.1 Destructive and constructive effect

As for our reconstruction of the argumentation structure of Searle's thought experimental report, in the starting p-context Searle raises the following p-problem:

(7) "[W]hat is it for one to mean something by what one says [...]?" (Searle 1969: 42)

The starting p-context includes Grice's (1989 [1957]) explication of the notion of 'non-natural meaning', which is one of the central components of his model of communication. Searle summarizes Grice's theory of meaning in the following thesis:¹⁶

(8) $0 < |G|$ "to say that a speaker *S* meant something by *X* is to say that *S* intended the utterance of *X* to produce some effect in a hearer *H* by means of the recognition of this intention." (Searle 1969: 43)

In the subscript, *G* stands for Grice's theory, which is the direct source of the plausibility value of (8). (8) receives a positive plausibility value on the basis of Grice (1989a [1957]) as a direct source. However, Searle seems to be ready to acknowledge Grice's paper as a reliable source only to a limited extent and only provisionally. The reason for this decision is the circumstance that Searle does not consider Grice's argumentation in support of (8) convincing, because, according to him,

(9) (8) "fails to account for the extent to which meaning can be a matter of rules or conventions". (Searle 1969: 43)

Obviously, Searle's thought experimental report is intended to have a destructive effect on Grice's theory of meaning in that it supports the claim in (9). Therefore, we now turn to the question of what the structure of a destructive thought experiment is.

Several classifications of thought experiments have been put forward. According to Brown and Fehige (2014), it has been generally accepted that in whatever way one defines thought experiments, two basic types are to be distinguished: destructive and constructive ones. Destructive thought experiments serve the rejection of theories, while constructive ones support theories.

¹⁶ Following Searle, we will speak simply of 'meaning', because we are not concerned with such uses of 'mean' as, for example, *Clouds mean rain*, which Grice (1989 [1957]) distinguishes from 'non-natural meaning'.

Peijnenburg and Atkinson (2003: 306) suggest the following basic inference pattern for destructive thought experiments:

- (10) Premises:
 (a) $(T \ \& \ E) \rightarrow Q$
 (b) $\sim Q$
 (c) E
 Conclusion:
 (d) $\sim T$

In Peijnenburg and Atkinson's account, T is a theory, E is a thought experiment, and Q is a particular situation of which it is generally known that it is not the case (that is, it is the case that $\sim Q$). Then, the pattern says that if the theory T is adjoined to the thought experiment E , then the theory is untenable. With respect to the refutation of T , the crucial element of this pattern is (10)(b). Referring to Parfit (1984), Peijnenburg and Atkinson (2003: 307) emphasize that $\sim Q$ is strong belief because everybody knows that Q is false: "a thought experiment can only be deemed successful if it induces the same – true or false – belief in the majority of people that are exposed to it".

We interpret the pattern in (10) in line with the following additional remarks:

First, T , E and Q are to be interpreted as statements. Trivially, to put it simply, theories are to be represented as statements with certain relations between them (that we do not go into here).

Second, from the p-model it follows that, contrary to Peijnenburg and Atkinson's assumption, the statements that the pattern consists of are not true or false, but rather, *plausible statements*. Namely, it goes without saying that the statements of the theory T are not true with certainty but merely have a plausibility value – otherwise the problem of whether it can be supported or rejected by a thought experiment would not arise. It is also clear that the statements constituting the thought experimental report cannot be assumed as certainly true, either.¹⁷

Relying on the pattern in (10), the destructive effect of Searle's thought experiment can be reconstructed as follows:¹⁸

- (11) Premises:
 (a) $0 < | \text{If to say that a speaker } S \text{ meant something by } X \text{ is to say that } S \text{ intended the utterance of } X \text{ to produce some effect in a hearer } H \text{ by means of the recognition of this intention, and in World War II Searle, as an American, captured by Italian soldiers, utters } \textit{Kennst du das Land, wo die Zitronen blühen?}, \text{ then Italian soldiers will come to believe that Searle is a German soldier.} |_{STE} < 1$
 (b) $0 < | \text{It is not the case that Italian soldiers will come to believe that Searle is a German soldier.} |_{STE} < 1$
 (c) $0 < | \text{In World War II Searle, as an American, captured by Italian soldiers, utters } \textit{Kennst du das Land, wo die Zitronen blühen?} |_{STE} < 1$
 Conclusion:

¹⁷ Kertész and Rákosi (2014b) presented a detailed analysis of Searle's thought experimental report on *Kennst du das Land, wo die Zitronen blühen?* in the framework of the p-model. Here we choose a much more abstract account in subsuming the core of the argumentation process under (11) and (14) below.

¹⁸ The inference pattern is *plausible modus tollens*: $0 < | \text{If } A, \text{ then } B |_S < 1$; $0 < | \sim B |_S < 1$; therefore, $0 < | \sim A |_S < 1$.

- (d) $0 < |I(11)| < 1$ It is not the case that to say that a speaker S meant something by X is to say that S intended the utterance of X to produce some effect in a hearer H by means of the recognition of this intention.

In (11)(a)-(c), STE stands for Searle's thought experiment as the direct source of the plausibility value of the statements at issue and $I(11)$ indicates that it is the inference in (11) that is the indirect source of the conclusion in (11)(d). Peijnenburg and Atkinson's remark about strong belief quoted above means, in the light of the p-model, that (11)(b) is to be assigned a high plausibility value and, as a result of this, the plausibility value of the conclusion, i.e. (11)(d), is assumed to be higher than that of (8).

Peijnenburg and Atkinson do not discuss the structure of constructive thought experiments. However, it is straightforward that if we accept the pattern in (10), then the structure of a constructive thought experiment will be that of *reduction* in Polya's sense.¹⁹ Accordingly, the pattern will be this:

- (12) Premises:
 (a) $(T \ \& \ E) \rightarrow Q$
 (b) Q
 (c) E
 Conclusion:
 (d) T

Most thought experiments are constructive and destructive at the same time (see e.g. Scott 2000, Brown and Fehige 2014): while they serve to refute one theory, they support another. Indeed, Searle's thought experiment is not only destructive with respect to Grice's theory of meaning, it is at the same time constructive as well. Namely, it is intended to assign a higher plausibility value to the plausible statement (13) than to its rival in (8):

- (13) $0 < |I(13)| < 1$ The analysis of illocutionary acts "must capture both intentional and conventional aspects of meaning and the relationship between them." (Searle 1969: 45)

Here, at the outset, Searle's task is to *find the source* that is capable of providing (13) with a plausibility value (see the question mark in the subscript). Then, the central plausible inference of his argumentation is (14).²⁰ This inference is the indirect source denoted by $I(14)$ in (14)(d) that assigns a plausibility value to (13).

- (14) Premises:
 (a) $0 < |I(13)| < 1$ If the analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them, and in World War II Searle, as an American, captured by Italian soldiers, utters *Kennst du das Land, wo die Zitronen blühen?*, then Italian Soldiers will not come to believe that Searle is a German soldier.
 (b) $0 < |I(13)| < 1$ Italian soldiers will not come to believe that Searle is a German soldier.
 (c) $0 < |I(13)| < 1$ In World War II Searle, as an American captured by Italian soldiers, utters *Kennst du das Land, wo die Zitronen blühen?*

¹⁹ Using our notation, the inference pattern of *reduction* is: $0 < |I(A)| < 1$; $0 < |B|_S < 1$; therefore, $0 < |A|_S < 1$. Polya evaluates reduction as "the simplest and most widespread pattern of plausible reasoning" (Polya 1948: 222). It is the plausible analogue to what Charles S. Peirce called 'abduction'. Nevertheless, the term 'reduction' goes back to Łukasiewicz (1970 [1912]: 7).

²⁰ The subscripts are analogous to those in (11).

Conclusion:

- (d) $0 < | \text{The analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them.} |_{I(14)} < 1$

Nevertheless, the plausible argumentation process carried out in the thought experiment does not terminate with the conclusion in (14)(d). Among others, the first reason for this is that it raised fierce debates. For example, Grice himself responded to it, and other authors such as Meggle and Ulkan (1979) reflected on it, too. The second reason is that, over the years, (14)(d) motivated a series of real experiments the task of which was to test this hypothesis. Some of these experiments assumed to have found evidence against it, and others intended to support it; further experiments continued the line of argumentation initiated by the thought experimental report.

After having clarified the constructive and the destructive effect of the thought experiment, in the next subsections we will proceed as follows. Subsection 3.2 will be devoted to the illustration of (RV)(a), i.e. arguments which claim that Searle's thought experiment yields contradictory conclusions. In subsection 3.3, we will exemplify the continuation of the thought experiment by real experiments in the light of (RV)(c). In Subsection 3.4, we will examine whether the thought experiment has experiential content and if so, in what sense (see also (RV)(d)).

3.2 On (RV)(a): Does Searle's thought experiment lead to contradictory conclusions?

3.2.1 A Wittgensteinian subphase of Searle's argumentation process

The argumentation that we have subsumed under (14) can be broken down into smaller argumentation steps. Within the elaboration of the p-context, one of them is this:²¹

- (15) "In the *Philosophical Investigations*, Wittgenstein (discussing a different problem) writes 'Say 'it's cold here' and *mean* 'it's warm here'". The reason we are unable to do this without further stage setting is that what we can mean is at least sometimes a function of what we are saying. Meaning is more than a matter of intention, it is also at least sometimes a matter of convention." Searle (1969: 45)

Here the p-context is *extended* by a new source that is a quotation from § 510 of Wittgenstein's *Philosophical Investigations*.²² The starting point of Searle's argumentation is the following datum (in the sense of the p-model):²³

- (16) $0 < | \text{Without further stage setting, when one says, } \textit{it's cold here}, \text{ one cannot mean } \textit{it's warm here.} |_W < 1$

Now, the following plausible inference presents itself for Searle:²⁴

²¹ For details, see Kertész and Rákosi (2014b).

²² The whole passage is this: "510. Make the following experiment: *say* 'It's cold here' and *mean* 'It's warm here'. Can you do it? – And what are you doing as you do it? And is there only one way of doing it?" (Wittgenstein 1953: 140; italics as in the original)

²³ At this point, Searle does not clarify what "further stage setting" means; he does this only later within the thought experiment. In (16), *W* stands for Wittgenstein's thought experimental report as the direct source of the plausibility of (16).

²⁴ The structure of the following inferences is *plausible modus ponens*: $0 < | \text{If } A, \text{ then } B |_S < 1$; $0 < | A |_S < 1$; therefore, $0 < | B |_S < 1$.

- (17) Premises:
- (a) $0 < |$ If without further stage setting, when one says, *it's cold here*, one cannot mean *it's warm here*, then there are cases in which what one can mean is a function of what one is saying. $|_{STE} < 1$
 - (b) $0 < |$ Without further stage setting, when one says, *it's cold here*, one cannot mean *it's warm here*. $|_W < 1$
- Conclusion:
- (c) $0 < |$ There are cases in which, what one can mean is a function of what one is saying. $|_{I(17)} < 1$

Our reconstruction of the Wittgensteinian situation quoted in (15) continues with a plausible inference the first premise of which is (17)(c):

- (18) Premises:
- (a) $0 < |$ There are cases in which, what one can mean is a function of what one is saying. $|_{I(17)} < 1$
 - (b) [$0 < |$ If there are cases in which what one can mean is a function of what we are saying, then meaning is more than a matter of intention, it is also at least sometimes a matter of convention. $|_{STE} < 1$]
- Conclusion:
- (c) $0 < |$ Meaning is more than a matter of intention, it is also at least sometimes a matter of convention. $|_{I(18)} < 1$

Due to the plausible argumentation process as summarized with the help of the above chain of inferences, at this point the p-context has been *retrospectively re-evaluated*. In particular, the source of the plausibility value of the datum in (16) is Wittgenstein's thought experimental report (symbolized by W) and, through the above inferences, it contributes to the decrease of the plausibility value of (8). Namely, apart from (14)(d), (18)(c) is also p-inconsistent with (8) (while (14)(d) and (18)(c) are of course consistent).

3.2.2 Meggle and Ulkan's counter-argumentation

In their well known paper, Meggle and Ulkan argued against the conclusions that Searle drew from his thought experiment. In Meggle and Ulkan's (1979) view, Searle's thought experiment aims to show that Grice's three conditions we quoted in (8) are necessary, but not sufficient for the explication of the notion of meaning. One of Searle's main arguments against Grice is that from Wittgenstein's claim that it is impossible to "[s]ay 'it's cold here' and mean 'it's warm here'", it follows that meaning is at least sometimes conventional (i.e. (18)(c)). Meggle and Ulkan's strategy is to question the plausibility of Wittgenstein's claim just quoted and to infer from this the implausibility of (18)(c), i.e. Searle's thesis, according to which "meaning is more than a matter of intention, it is also at least sometimes a matter of convention".

The first step in their argumentation is to consider different uses of the term 'mean' in Grice's, Wittgenstein's and Searle's writings cited. Meggle and Ulkan interpret Wittgenstein's notion of meaning as

- (19) 'by uttering x somebody wants to make somebody else understand something'.

However, in this case, Wittgenstein's thesis, according to which meaning is also a matter of convention, turns out to be false. Namely, although the utterance x may have no conventional meaning, it is possible that by uttering x somebody wants to make somebody else understand

something (Meggles and Ulkan 1979: 79). For example, a person is in the house and he whistles so as to make his father, who is in the garden, understand that the visitor they were waiting for has arrived (Meggles and Ulkan's example B-1). However, the whistle has no conventional meaning. That is, Wittgenstein specified (19) to

- (20) 'by uttering a conventional expression of a language x somebody wants to make somebody else understand something'.

This boils down to the trivial claim that "conventional communication is possible only if there exist expressions with a conventional meaning" (Meggles & Ulkan 1979: 80) and it is exactly this what Searle (according to Meggles and Ulkan) claims, too.

In Meggles and Ulkan's opinion, demonstrating this is sufficient to undermine both Wittgenstein's account of meaning and its extension to Searle's approach. Therefore, Searle's thought experiment seems to have failed, too, because it does not affect Grice's theory of meaning in (8). Namely, Meggles and Ulkan conclude that in Wittgenstein's and in Searle's view the conventional nature of meaning is not restricted to *how* one communicates, but involves the *content* of communication, too. Consequently, while Grice attempted to explicate the notion of ' S means that p by uttering A ' which abstracts from the content of the utterance at the outset, Searle (and Wittgenstein) focused on the considerably narrower notion of ' S conventionally means that p by uttering A ' which involves the content of the utterance.

In a very simplified way, the core of the plausible argumentation process Meggles and Ulkan carry out can be reconstructed as in (21)-(23):²⁵

(21) Premises:

- (a) $0 < |$ If Wittgenstein's theory of meaning is to be interpreted as 'by uttering x somebody wants to make somebody else understand something', then Wittgenstein's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false. $|_{MU} < 1$
- (b) $0 < |$ Wittgenstein's theory of meaning is to be interpreted as 'by uttering x somebody wants to make somebody else understand something'. $|_{MU} < 1$

Conclusion

- (c) $0 < |$ Wittgenstein's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false. $|_{I(21)} < 1$

(22) Premises

- (a) $0 < |$ If Wittgenstein's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false, then Searle's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false, too. $|_{MU} < 1$
- (b) $0 < |$ Wittgenstein's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false. $|_{I(21)} < 1$

Conclusion:

- (c) $0 < |$ Searle's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false $|_{I(22)} < 1$

(23) Premises

²⁵ In the following three inferences the subscript MU refers to Meggles and Ulkan's paper as a direct source of the plausibility of the statements at issue.

- (a) $0 < | \text{If Searle's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false, then Searle's thought experiment failed.} |_{MU} < 1$
 - (b) $0 < | \text{Searle's thesis according to which meaning is more than a matter of intention, it is also at least sometimes a matter of convention, is false.} |_{I(22)} < 1$
- Conclusion:
- (c) $0 < | \text{Searle's thought experiment failed.} |_{I(23)} < 1$

If we accept Meggle and Ulkan's reasoning, then Searle's thought experiment seems to fail in the sense of (RV)(a), for *it gives rise to contradictory conclusions*, namely, (18)(c) and (22)(c).

However, Meggle and Ulkan also maintain that, although, according to the above line of argumentation, Searle's thought experiment does not affect Grice's theory of meaning, it also yields a correct conclusion:

- (24) Premises:
- (a) $0 < | \text{If the claim 'meaning is more than a matter of intention, it is also at least sometimes a matter of convention' is about conventional communication, then meaning is more than a matter of intention, it is also at least sometimes a matter of convention.} |_{MU} < 1$
 - (b) $0 < | \text{The claim 'meaning is more than a matter of intention, it is also at least sometimes a matter of convention' is about conventional communication'.} |_{MU} < 1$
- Conclusion:
- (c) $0 < | \text{Meaning is more than a matter of intention, it is also at least sometimes a matter of convention.} |_{I(24)} < 1$

Meggle and Ulkan emphasize that here 'meaning' is a different notion than in (8), and therefore the conclusion in (24)(c) is plausible (but trivial). Thus, our analysis of Meggle and Ulkan's argumentation suggests that both (24)(c) and (8) can be maintained and that both of them are statements with a high plausibility value.

3.2.3 Conclusions

Meggle and Ulkan's finding according to which Grice and Searle tackle different notions of meaning clearly exemplifies the basic properties of the process of plausible argumentation in which the thought experiment participates. Namely, first, Searle's thought experiment starts a cycle of the retrospective re-evaluation of Grice's explication of meaning. This cycle is also prismatic, because through the prism of the thought experiment, Grice's explication is viewed from a new perspective. Second, similarly, within an analogous process, Searle re-evaluates Wittgenstein's statement. What is more, third, Meggle and Ulkan's reasoning from the thought experiment toward their own conclusion constitutes a further cycle of plausible argumentation which retrospectively re-evaluates both Grice's and Searle's claims. This retrospective re-evaluation is prismatic, too, because it is carried out from a new perspective, again. Consequently, it is not the case that Searle's thought experiment yielded contradictory results, but rather, it triggered a process of plausible argumentation that has led to a new finding, namely, the distinction between two notions of 'meaning' and their compatibility. Therefore, instead of being a poor one, in this sense Searle's thought experiment turned out to be effective. Thus, our reconstruction of the above phases of the thought experiment, as well as Meggle and Ulkan's counter arguments, support our hypothesis (H).

At this point we interrupt the reconstruction of the argumentation structure of the thought experimental report. In the next section we turn to that of real experimental reports that are closely related to what we have discussed so far.

3.3 On (RV)(c): Does Searle's thought experiment interact with real experiments?

3.3.1 The starting p-context

According to (RV)(c), one indication of poor thought experiments is that they are not connected to real experiments. Indeed, by the time Searle's thought experimental report appeared, it did not motivate real experiments that could have led to further findings starting from (14)(d), which is the constructive result of Searle's thought experiment. Nevertheless, over the decades this situation has changed. Among many other attempts, Holtgraves (2008) and Holtgraves and Ashley (2001) reported on a series of experiments in which the main finding of Searle's thought experiment in (14)(d) was tested. In the present subsection we will exemplify the way in which (14)(d) interacts with real experiments within a complex process of plausible argumentation. Thereby, as the continuation of Kertész and Rákosi (2014b) that focused on Holtgraves and Ashley (2001), we will analyse the experimental reports in Holtgraves (2008) in order to illustrate our hypothesis (H).

In the starting p-context, Holtgraves presupposes (14)(d), which he reformulates as the following plausible statement:

- (25) 0<|"It is the illocutionary act that most closely captures the nature of the speaker's intention in producing a particular conversation turn."|_{STE}<1 (Holtgraves 2008: 362)

He adds the following p-problem to the final p-context of Searle's thought experiment:

- (26) What is the exact nature of intention recognition in illocutionary acts?

In order to answer this question by using real experiments, in the starting p-context the author contrasts two rival hypotheses. One is that of relevance theory:

- (27) "Following Grice (1975), Sperber and Wilson assume that each conversational utterance carries with it a presumption of relevance. In general, the presumption of relevance suggests that comprehenders will search for the first interpretation of an utterance (least effort) that yields maximum relevance. In this theory, illocutionary force recognition is not required for utterance comprehension. Hence, *I will definitely do it tomorrow* does not need to be categorized as a promise in order to be understood." (Holtgraves 2008: 362)

The quotation in (27) is based on the following inference:

- (28) Premises:
- (a) 0<|If the presumption of relevance suggests that comprehenders will search for the first interpretation of an utterance (least effort) that yields maximum relevance, then illocutionary force recognition is not required for utterance comprehension.|_{G&RT}<1
 - (b) 0<|The presumption of relevance suggests that comprehenders will search for the first interpretation of an utterance (least effort) that yields maximum relevance.|_{G&RT}<1
- Conclusion:
- (c) 0<|Illocutionary force recognition is not required for utterance comprehension.|_{I(28)}<1

(28)(c) emerges as the consequence of the retrospective re-evaluation of Grice's theory of meaning which (8) is a part of within the complex process of plausible argumentation.

Accordingly, *RT* refers to this source of the plausibility of (28)(a) and (b), and *G* refers to Grice (1989 [1975]) as another direct source. In this process of plausible argumentation, (28)(c) has been obtained via the continuation of Grice's thought experiments in that the proponents of relevance theory retrospectively re-evaluated them. At this point we do not go into an analysis of this process, we just notice (28).²⁶

Another possible solution to the p-problem raised in (26) presents itself as the continuation of the argumentation process which was initiated by Searle's thought experiment and which provisionally yielded (14)(d) (re-formulated as (25)) with a high plausibility value. According to Holtgraves (2008: 362), "[...] speech act theory suggests that illocutionary force recognition plays a critical role in the comprehension of conversation remarks". We reconstruct the corresponding inference in (29):

(29) Premises:

(a) $0 < | \text{If it is the illocutionary act that most closely captures the nature of the speaker's intention in producing a particular conversation turn (see (25)), then illocutionary force recognition plays a critical role in the comprehension of conversation remarks.} |_H < 1$

(b) $0 < | \text{It is the illocutionary act that most closely captures the nature of the speaker's intention in producing a particular conversation turn.} |_{STE \& (25)} < 1$

Conclusion:

(c) $0 < | \text{Illocutionary force recognition plays a critical role in the comprehension of conversation remarks.} |_{I(29)} < 1$

The subscript *H* stands for Holtgraves' paper as the direct source of the plausibility of the statements in (29)(a). (28)(c) and (29)(c) are *rival hypotheses*. Since each of them is the negation of the other, the starting p-context of Holtgraves' argumentation is informationally overdetermined in the sense of the p-model and is thus *p-inconsistent*. Accordingly, real experiments may be expected to resolve this inconsistency by modifying the p-context in such a way that (29)(c) receives a higher plausibility value than (28)(c). Indeed, Holtgraves and Ashley (2001) reported on real experiments that increased the plausibility of (29)(c). Then, the comparison of the plausibility values of (28)(c) and (29)(c) yields the following evaluation:

(30) $0 < | \text{Illocutionary force recognition is not required for utterance comprehension.} |_{RT} < | \text{Illocutionary force recognition plays a critical role in the comprehension of conversation remarks.} |_{STE \& HA} < 1$

HA stands for Holtgraves and Ashley (2001).²⁷

3.3.2 First phase of the argumentation cycle: The new p-context-version

In Holtgraves (2008), the starting p-context is enriched by a new p-problem raised in the following quotation:

(31) "What is not clear, however, is whether illocutionary force also forms part of the long-

²⁶ See Kertész and Kiefer (2013) for discussion.

²⁷ Analogously to what we remarked with respect to the debate between Grice and Searle, here we also add that it is not our intention to take sides between the standpoints of relevance theory and that of Holtgraves' claims. Rather, we just present a partial metatheoretical reconstruction of the argumentation structure of Holtgraves' (2008) experimental report.

term representation of an utterance. Does Andy's representation of Bob's utterance *I'll definitely do it tomorrow* include the illocutionary force of the utterance — that it's a promise? It is possible that illocutionary force activation will play a role during comprehension, but is then quickly forgotten, similar to the manner in which people quickly forget the exact wording of an utterance. Alternatively, it is possible that illocutionary force activation will play a role in comprehension and subsequently form part of the long-term representation of that utterance. In speech act theory, most illocutionary points have the logical form $F(p)$, where F represents the illocutionary force and p represents the propositional content (Searle & Vanderveken, 1985; Vanderveken, 1990). In this view, the long-term representation of the utterance *I'll definitely do it tomorrow* would be something like *promise* (will perform the act tomorrow)." Holtgraves (2008: 362-363)

Thus, the p-problem is this:

- (32) Does illocutionary force also form part of the long-term representation of an utterance?

As the quotation suggests, two rival hypotheses present themselves as possible solutions to (32):

- (33) $0 < | \text{Illocutionary force also forms part of the long-term representation of an utterance.} |_{STE\&H} < 1$
- (34) $0 < | \text{Illocutionary force does not form part of the long-term representation of an utterance.} |_{G\&RT} < 1$

Obviously, the p-context is informationally overdetermined, that is, it is *p-inconsistent*. Consequently, the aim of the next phases of the argumentation process is to find a solution to the p-problem raised in (32) by the modification of the p-context so that it will become less overdetermined, i.e. consistent through the decision between (33) and (34). The means by which the author wants to achieve this is the extension of the p-context by five real experiments. In what follows, we will reconstruct the core of the argumentation process involving the real experimental reports without going into the details.

Holtgraves' central idea is that if illocutionary force activation is part of the long-term representation of the utterance, then at least two consequences suggest themselves (Holtgraves 2008: 363). The first is that people will falsely remember conversation utterances as containing speech act verbs characterizing those utterances. The second says that speech act verbs will serve as recall cues for utterances that perform the speech acts at issue. Experiments 1 through 3 were devoted to the first consequence, and 4 and 5 to the second.

3.3.3 Second phase of the argumentation cycle: The elaboration of the p-context

3.3.3.1 Subcycle 1

In experiments 1 and 2 – which correspond to the first subcycle of the elaboration of the p-context – the participants were presented two sets of scenarios. Those in the first set described a situation between two persons and were followed by a remark which was made by one of these persons and which performed an illocutionary act. The second set was the control set. In the control scenarios, the utterances shared as many words as possible with the speech act scenarios but the final utterance did not perform the relevant illocutionary act.

The core of the argumentation governing the first two experiments can be reconstructed as the following plausible inference with the structure of *reduction* (Holtgraves 2008: 364):

- (35) Premises:
- (a) $0 < | \text{If illocutionary force is retained in long-term memory, then participants are more likely to falsely recognize a lure containing a relevant speech act verb after reading an utterance that performs that speech act than after reading an utterance that does not perform that speech act.} |_{RE1\&2} < 1$
 - (b) $0 < | \text{Participants are more likely to falsely recognize a lure containing a relevant speech act verb after reading an utterance that performs that speech act than after reading an utterance that does not perform that speech act.} |_{RE1\&2} < 1$
- Conclusion:
- (c) $0 < | \text{Illocutionary force is retained in long-term memory.} |_{I(35)} < 1$

In the subscripts, *RE1&2* stands for the first two real experiments as direct sources. The difference between the two experiments was merely that while in Experiment 1 the target was always a description of a conversation utterance (such as *Bob said that he was right, that it's wrong to experiment on animals*), in Experiment 2, the target was the conversation utterance (e.g. *That's right, it's wrong to experiment on animals*).

The findings which the two experiments yielded supported the plausibility of (35)(b) and via this, that of (35)(c).

The information content of (35)(c) is equivalent to that of (33), but – due to the indirect source represented in the inference (35) – the plausibility value of the former is higher than the plausibility value of both of (33) and (34). Thus, (35)(c) provides a provisional solution to the p-problem raised in (32).

3.3.3 Subcycle 2

Then, a new subcycle is started which corresponds to Experiment 3. The p-context is modified by the introduction of new data the aim of which is to perform a recall test. The reason for this extension is that recognition tests are sensitive to false memory effects. Here again, the argumentation process associated with the experiment consists of a series of interconnected plausible inferences which can be subsumed under (36)-(38). The inference in (36) has the structure of *reduction*:

- (36) Premises:
- (a) $0 < | \text{If conversation utterances are tagged with the relevant speech act verbs, then participants are more likely to falsely recall the speech act verb after having read the speech act version than after having read the control version.} |_{RE3} < 1$
 - (b) $0 < | \text{Participants are more likely to falsely recall the speech act verb after having read the speech act version than after having read the control version.} |_{RE3} < 1$
- Conclusion:
- (c) $0 < | \text{Conversation utterances are tagged with the relevant speech act verbs.} |_{I(36)} < 1$

RE3 stands for Experiment 3 as a direct source. The structure of the next inference is *plausible modus ponens*:

- (37) Premises:
- (a) $0 < | \text{If conversation utterances are tagged with the relevant speech act verbs, then illocutionary force is retained in long-term memory.} |_{RE3} < 1$
 - (b) $0 < | \text{Conversation utterances are tagged with the relevant speech act verbs.} |_{I(36)} < 1$
- Conclusion:
- (c) $0 < | \text{Illocutionary force is retained in long-term memory.} |_{I(37)} < 1$

Here the experiment focused on producing data that increase the plausibility of (37)(b). Thus, the experiment related to (37) serves as another direct source to increase the plausibility of the same thesis that has already been obtained as (35)(c) (see also the hypothesis in (33)).

Then, a *reduction* follows, again:

- (38) Premises:
- (a) $0 < | \text{If speech act verbs serve as quick indicators of a speaker's intention, then they are more likely to be recalled when people are trying to recall the gist of what someone has said.} |_{RE3} < 1$
 - (b) $0 < | \text{Speech act verbs are more likely to be recalled when people are trying to recall the gist of what someone has said.} |_{RE3} < 1$
- Conclusion:
- (c) $0 < | \text{Speech act verbs serve as quick indicators of a speaker's intention.} |_{I(38)} < 1$

3.3.3.3 Subcycle 3

The p-context is further extended by the next subcycle of the argumentation process, focusing on Experiment 4. The aim of Experiment 4 was to test the finding that speech act verbs are stored as utterance tags influencing both the recognition and the reconstruction of the utterance that has to be remembered. The series of inferences constituting this new subcycle can be subsumed under (39), which has the structure of *reduction*:

- (39) Premises:
- (a) $0 < | \text{If the speech act performed with an utterance forms part of the representation of the utterance, then the speech act verb serves as a good retrieval cue for the utterance.} |_{RE3} < 1$
 - (b) $0 < | \text{The speech act verb serves as a good retrieval cue for the utterance.} |_{RE3} < 1$
- Conclusion:
- (c) $0 < | \text{The speech act performed with an utterance forms part of the representation of the utterance.} |_{I(39)} < 1$

The data which the experiment yielded supported the plausibility of (39)(b) and via this, that of (39)(c).

3.3.3.4 Subcycle 4

The last experiment with which a new subcycle of the argumentation process started introduced additional information into the p-context. The source of the plausibility of this new information is Holtgraves (1997), in which the Conversation Indirectness Scale was developed. The Conversation Indirectness Scale measures, among other things, the *interpretation dimension* – that is the extent to which a person looks for nonliteral meanings

in the utterances of the participants of a conversation (Holtgraves 2008: 367). The p-context of the subcycle focusing on Experiment 5 is then supplemented by the following claim rooted in Holtgraves (1997):

- (40) "People scoring high on this dimension are relatively more likely to interpret others' utterances as conveying nonliteral meanings, and they are faster at recognizing these meanings [...]. Illocutionary force represents an inference of sorts; it is not literally present in the utterance. People who score high on the interpretation dimension, and hence look for nonliteral meanings, should tend to characterize utterances in terms of their illocutionary force." Holtgraves (2008: 367)

Then, the main inference which the experiment is based on presents itself as the *reduction* in (41):

- (41) Premises:
- (a) $0 < | \text{If speech act words are significantly better cues for people who score high on the interpretation dimension of the Conversation Indirectness Scale, then these people tend to characterize utterances in terms of their illocutionary force.} |_{RE5\&(40)} < 1$
 - (b) $0 < | \text{People who score high on the interpretation dimension of the Conversation Indirectness Scale tend to characterize utterances in terms of their illocutionary force.} |_{I(40)} < 1$
- Conclusion:
- (c) $0 < | \text{Speech act words are significantly better cues for people who score high on the interpretation dimension of the Conversation Indirectness Scale.} |_{I(41)} < 1$

3.3.4 Third phase of the argumentation cycle: The comparison of the p-context versions

At this point of the argumentation process, the rival p-context versions can be compared. We reconstruct this as follows:²⁸

- (42) Premises:
- (a) $0 < | \text{If (35)(c) and (36)(c) and (37)(c) and (38)(c) and (39)(c) and (41)(c), then (33).} |_{RE1-5} < 1$
 - (b) $0 < | \text{(35)(c) and (36)(c) and (37)(c) and (38)(c) and (39)(c) and (41)(c).} |_{RE1-5} < 1$
- Conclusion:
- (c) $0 < | \text{(33)} |_{I(42)} < 1$

Thus we obtain that, as a result of the real experiments, the plausibility value of the hypothesis in (33) is higher than that of (34):

- (43) $0 < | \text{Illocutionary force does not form part of the long-term representation of an utterance.} |_{I(34)} < | \text{Illocutionary force also forms part of the long-term representation of an utterance.} |_{I(42)} < 1$

²⁸ For the sake of transparency, here we refer to the corresponding plausible statements by their number, instead of quoting them.

In this way, through the real experiments 1-5, (33) and (34) have been retrospectively re-evaluated in the course of the plausible argumentation process.

3.3.5 The final p-context

The final p-context retrospectively re-evaluates both the result of Searle's thought experiment i.e. (14)(d) and – via reference to relevance theory – Grice's original theory of meaning as quoted in (8) (see also the quotation in (27) as well):

- (44) "Although there is wide agreement regarding the importance of intention recognition in conversation processing, there is disagreement on the nature of that recognition. Fundamental to speech act theory (Searle, 1969) is the notion that people categorize utterances in terms of the speech act performed; that is, recognition of illocutionary force is entailed in comprehension. Note that speech act recognition satisfies both of the above requirements; it captures with a single word (good-enough processing) the intended action performed by the speaker. Relevance theory (Sperber & Wilson, 1995) and certain computation models explicitly deny the necessity of this step. In this view, speech act categorization is possible after the fact, but is not a necessary component of comprehension." Holtgraves (2008: 369)

As the above quotation witnesses, the finding of the real experiments – i.e. the circumstance that the hypothesis (33) receives a high plausibility value – retrospectively re-evaluates (14)(d), which was the main thesis Searle's thought experiment yielded. This argumentation process can be subsumed under the *reduction* in (45):

- (45) Premises:
- (a) $0 < \text{[If the analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them (see (14)(d) and (25)), then Illocutionary force also forms part of the long-term representation of an utterance (see (33)).]}_{RE1-5} < 1$
 - (b) $0 < \text{[Illocutionary force also forms part of the long-term representation of an utterance.]}_{RE1-5} < 1$
- Conclusion:
- (c) $0 < \text{[The analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them.]}_{I(45)} < 1$

(33) supported (14)(d), and (14)(d) was the rival to (8) (i.e. to Grice's thesis). Consequently, at this stage of the argumentation the plausibility of (14)(d) is rooted in at least two sources: Searle's thought experiment and Holtgraves' real experiments. Thus, these two sources jointly assign a higher plausibility value to (14)(d) than to (8):²⁹

- (46) $0 < \text{[To say that a speaker } S \text{ meant something by } X \text{ is to say that } S \text{ intended the utterance of } X \text{ to produce some effect in a hearer } H \text{ by means of the recognition of this intention (see (8)).]}_G < \text{[The analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them (see (14)(d)).]}_{TE\&RE1-5} < 1$

²⁹ (46) must not be interpreted in such a way that we take sides for Searle against Grice. This is merely the result of the reconstruction of Searle's and Holtgraves' lines of argumentation, as already mentioned, and *not our* decision.

3.3.5 Conclusions

As we have seen, there is a clear-cut feedback mechanism between the thought experiment and the real experiments exemplified here. On the one hand, as we have just seen, the main finding of the thought experiment (i.e. (14)(d)) is part of the starting p-context of the real experiments. Thus, it motivates and shapes the subsequent process of plausible argumentation inherent in the experimental report. On the other hand, as (45) witnesses, the result that the real experiments yielded through a cyclic return prismatically and retrospectively increased the plausibility value of the thought experiment's finding and contributed to the possible decision between the latter and its rival.

3.4 On (RV)(d): Does Searle's thought experiment have experiential content?

There are at least two reasons why it is well motivated to turn to the question of whether Searle's thought experiment has experiential content and if so, of what kind.

First, according to the literature on thought experiments, one of the main problems is how it can be the case that "[...] thought experiments often have *novel empirical import* even though they are conducted entirely inside one's *head* [...]" (Horowitz & Massey 1991: 1; emphasis added). The standard answer to this question says that thought experiments would not work if some kind of empirical content did not enter them at the outset. So, it is only natural that (RV)(d) says that a thought experiment is poor if it lacks empirical content. In the previous subsection we have seen that Searle's thought experiment can be related to experiences via Holtgraves' real experiments and – as opposed to what one may expect on the basis of its philosophical nature – it does not fail in this respect. However, this kind of empirical content is only indirectly and retrospectively related to the thought experiment. Thus, it should be examined whether there is some kind of other experiential information directly included in the thought experiment.

Second, in the literature there is also agreement on the fact that thought experiments are counterfactual answers to 'What if...?' questions (see e.g. Cooper 2005, Rescher 2005 etc.). The question is: What would be the case if in World War II, Searle, as an American, were captured by Italian soldiers, and uttered *Kennst du das Land, wo die Zitronen blühen?* Accordingly, Searle's thought experiment tells an imaginary story in order to give a counterfactual answer to this question. Since the answer is counterfactual and the situation is imaginary, it need not have experiential content at the outset.³⁰

Nevertheless, the thought experiment itself includes a special kind of empirical content. While the real experiments obviously include perceptual data collected by the experimenter, in Searle's thought experimental report there are no data whose plausibility is rooted in direct contact with the perception of the empirical world as a direct source.³¹ In spite of this, the p-context of Searle's thought experimental report is not void of experiential data that arise from Searle's former everyday experiences with real-world situations analogous to

³⁰ It is this aspect which explains why in the Appendix we highlighted those expressions that indicate the hypothetical nature of the situation described in the thought experimental report.

³¹ In the light of the p-model, perceptual data are plausible statements the plausibility value of which is rooted in perception as a direct source. Perception is understood in accordance with this notion used in the philosophy of science, i.e. it includes not only human sensual perception but measurement results obtained by the use of technical instruments that are accepted as reliable to a certain extent in the scientific community, too.

the imaginary situation constructed in the thought experiment. For example, the starting p-context includes the following data (in the sense of the p-model):

- (47) (a) 0<|In the imaginary situation that the thought experiment is about, John Searle says, *Kennst du das Land...* etc. because he wishes to get Italian soldiers to believe that he is a German soldier._{|EE<1}
- (b) 0<|In the imaginary situation that the thought experiment is about, Italian soldiers do not know what the utterance *Kennst du das Land...* etc. means._{|EE<1}
- (c) 0<|In the imaginary situation that the thought experiment is about, Italian soldiers can realise that the utterance *Kennst du das Land...* etc. is a German sentence._{|EE<1}

Here *EE* stands for the direct source of the plausibility of the data in (47). It refers to a series of everyday experiences Searle had made earlier. The data in (47) contain only everyday concepts and no theoretical ones. In a further subcycle of the argumentation process presented in the thought experimental report, Searle retrospectively re-evaluates the statements in (47) through the prism of his theory of meaning insofar as the everyday concepts are replaced by the notions of his theory of meaning. Accordingly, the experiential data in (47) are constitutive components of the cyclic and prismatic process of plausible argumentation as manifested in the thought experimental report.

Another example is the following datum, which is part of the starting p-context of Searle's argumentation against (8) (see also Kertész and Rákosi 2014b):³²

- (48) 0<|According to the rules of German, as well as what Searle remembers, the words *Kennst du das Land...* etc. mean that 'Knowest thou the land where the lemon trees bloom?'._{|EE&KG<1}

The extension of the p-context by (48) makes it possible to draw the following plausible inference:

- (49) Premises:
- (a) 0<|If according to the rules of German, as well as what Searle remembers, the words *Kennst du das Land...* etc. mean 'Knowest thou the land where the lemon trees bloom?', then in imagining the situation depicted there is no difference between the meaning of these words according to the rules and conventions of the German language as Searle remembers it and what Searle means with them in the given situation._{|TE<1}
- (b) [0<|According to the rules of German, as well as what Searle remembers, the words *Kennst du das Land...* etc. mean 'Knowest thou the land where the lemon trees bloom?'._{|EE<1}]
- Conclusion:
- (c) 0<|In the imaginary situation depicted there is no difference between the meaning of these words according to the rules and conventions of the German language as Searle remembers it and what Searle means with them in the given situation._{|I(49)<1}

To conclude, although Searle's thought experimental report does not contain perceptual data, it includes a special kind of everyday experiential data. By their integration into the p-context, they are constitutive components of the plausible argumentation process which cyclically,

³² In the subscript *KG* refers to Searle's knowledge of German.

prismatically and retrospectively re-evaluate the pieces of information eventually leading to the plausibility of Searle's claim as represented in (14)(d). Therefore, in this respect, the thought experiment is not poor, either.

4 Summary

In the present paper we applied Kertész and Rákosi's p-model to the problem (P). The p-model yielded the hypothesis (H) as a possible solution which we contrasted with the Received View as summarized in (RV). Then, we presented a case study the aim of which was to exemplify the tenability of (H).

(H) seems to have shed fresh light on the nature of thought experiments in pragmatics. If we accept (H), then the following general insights suggest themselves:

(a) The philosophical literature treats thought experiments as deductive or inductive inferences (see e.g. Norton 2004), as vistas in a Platonic world (Brown 1991), as mental models (Nercessian 1992, Mišćević 1992), as models of whatever kind (Cooper 2005), or as several other kinds of things. Our approach differs from all accounts discussed in the literature in that it conceives of the thought experimental report we analysed as a cyclic and prismatic phase in a dynamic process of plausible argumentation. Thereby, the main finding which a thought experimental report contains increases the plausibility value of particular hypotheses and decreases that of others. None of the approaches known so far focuses on the notion of plausibility.

(b) (H) is in agreement with Sorensen (1992) as cited in the Introduction, according to which the crucial aspect of the distinction between good and poor thought experiments is whether the thought experiment is effective. However, (H) goes beyond Sorensen's claim in that it relates heuristic effectiveness to the plausible argumentation process. Then, effectiveness means that the thought experiment is capable of accomplishing its heuristic aim, namely, of governing the plausible argumentation process in such a way that the latter leads to a final p-context which includes a (provisional) solution or resolution of the p-problem the thought experiment at issue has tackled. This is a different answer to the question asked in (P) than the enumeration of the criteria in (RV).

(c) We have seen that a philosophical thought experiment that should be deemed poor if (RV) were feasible turned out to be fruitful. Consequently, in contrast to Peijnenburg and Atkinson's (2003) claim, it does not matter whether the thought experiment at issue is of a philosophical or of a scientific origin. What matters with respect to its effectiveness is solely whether and if so, to what extent, the thought experiment conforms to the process described in (H).

In particular, with respect to the comparison of (H) to (RV)(a), (c) and (d), the main insights can be summarized as follows:

(a) If (H) holds, then the emergence of the kind of contradiction mentioned in Peijnenburg and Atkinson (2003) does not make a thought experiment automatically poor. In particular, as we have seen, the plausible argumentation process can overrule the destructive consequences of contradictory conclusions drawn from the same thought experiment by different people during specific phases of a complex argumentation process (see Section 3.2. on (RV)(a)).

(b) In addition, if a thought experiment in pragmatics is seen as part of a process of plausible argumentation, then the relation between the thought experiment and the corresponding real experiments can be revealed in a unified way. This relation is also based on the feedback mechanism of the cyclic, prismatic and retrospective re-evaluation of information (Section 3.3). Thus, the relation between (H) and (RV)(c) is different from that between (H) and (RV)(a), because (RV)(c) is subsumed under (H).

(c) It is the same mechanism that facilitates the inclusion of relevant experiential information – such as personal experiences and/or information obtained from real experiments – in the thought experiment that contributes to the findings it is intended to lead to (Section 3.4 and (RV)(d)).

References

- Brown, James Robert. 1991. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London and New York: Routledge.
- Brown, James Robert and Yiftach Fehige. 2014. Thought experiments. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/thought-experiment/>. Accessed 22th July 2015.
- Buzzoni, Marco. 2008. *Thought Experiment in the Natural Sciences: An Operational and Reflexive-Transcendental Conception*. Würzburg: Königshausen and Neumann.
- Camilleri, Kristian. 2014. Toward a constructivist epistemology of thought experiments in science. *Synthese* 191: 1697–1716.
- Cohnitz, Daniel. 2006a. *Gedankenexperimente in der Philosophie*. Paderborn: Mentis.
- Cohnitz, Daniel. 2006b. Poor thought experiments? A comment on Peijnenburg and Atkinson. *Journal for General Philosophy of Science* 37: 373–392.
- Cooper, Rachel. 2005. Thought experiments. *Metaphilosophy* 36: 328-347.
- Gendler, Tamar Szabó. 2000. *Thought Experiments: On the Power and Limits of Imaginary Cases*. New York: Garland Publishing.
- Grice, Paul. 1989a [1957]. Meaning. In *In Studies in the Way of Words*, by Paul Grice, 213-223. Cambridge, MA.: Harvard University Press.
- Holtgraves, Thomas. 1997. Styles of language use: Individual and cultural variability in conversational indirectness. *Journal of Personality & Social Psychology* 73: 624-637.
- Holtgraves, Thomas. 2008. Conversation, speech acts, and memory. *Memory & Cognition* 36: 361-374.
- Holtgraves, Thomas and Aaron Ashley. 2001. Comprehending illocutionary force. *Memory & Cognition* 29: 83-90.
- Horowitz, Tamara and Gerald J. Massey. 1991. Introduction. In *Thought Experiments in Science and Philosophy*, ed. by Tamara Horowitz and Gerald J. Massey, 1-28. Savage MD: Rowman and Littlefield.
- Kertész, András. 2015. The puzzle of thought experiments in conceptual metaphor research. *Foundations of Science* 20: 147-174.
- Kertész, András and Ferenc Kiefer. 2013. From thought experiments to real experiments in pragmatics. In *Perspectives on Philosophy and Pragmatics*, ed. by Alessandro Capone, Franco Lo Piparo and Marco Carapezza, 53-86. Cham, Heidelberg, New York, Dordrecht and London: Springer.
- Kertész, András and Csilla Rákosi. 2009. Cyclic vs. circular argumentation in the Conceptual Metaphor Theory. *Cognitive Linguistics* 20: 703-732.
- Kertész, András and Csilla Rákosi. 2012. *Data and Evidence in Linguistics: A Plausible Argumentation Model*. Cambridge: Cambridge University Press.
- Kertész, András and Csilla Rákosi. 2013. Paraconsistency and Plausible Argumentation in Generative Grammar: A Case Study. *Journal of Logic, Language and Information* 22: 195-230.
- Kertész, András and Csilla Rákosi. 2014a. The P-model of data and evidence in linguistics. In András Kertész and Csilla Rákosi eds., 2014, 15-48.

- Kertész, András and Csilla Rákosi. 2014b. Thought experiments and real experiments as converging data sources in pragmatics. In András Kertész and Csilla Rákosi eds., 2014, 221-269.
- Kertész, András and Csilla Rákosi eds. 2014. *The Evidential Basis of Linguistic Argumentation*. Amsterdam and Philadelphia: Benjamins.
- Kertész, András and Csilla Rákosi. 2016. On the inferential structure of indirect reports. In *Indirect Reports and Pragmatics: Interdisciplinary Studies*, ed. by Alessandro Capone, Ferenc Kiefer and Franco Lo Piparo, 435-470. Cham, Heidelberg, New York, Dordrecht and London: Springer.
- Kühne, Ulrich. 2005. *Die Methode des Gedankenexperiments*. Frankfurt am Main: Suhrkamp.
- Łukasiewicz, Jan. 1970 [1912]. Creative elements in science. In *Selected Works*, by Jan Łukasiewicz, 12-44. Amsterdam: North Holland.
- Meggle, Georg and Maria Ulkan. 1979. Kennst du das Land, wo die Zitronen blühen? (Zu Searles Kritik and Grice). *Papiere zur Linguistik* 20: 75-87.
- Mey, Jacob L. 1987. Poet and peasant: A pragmatic comedy in five acts. *Journal of Pragmatics* 11: 281-297.
- Mey, Jacob L. 1993. *Pragmatics: An Introduction*. Oxford: Blackwell.
- Miščević, Nenad. 1992. Mental models and thought experiments. *International Studies in the Philosophy of Science* 6: 215-226.
- Moue, Aspasia S., Kyriakos A. Masavetas and Haido Karayianni. 2006. Tracing the developments of thought experiments in the philosophy of natural sciences. *Journal for the General Philosophy of Science* 37: 61-75.
- Nersessian, Nancy. 1993. In the theoretician's laboratory: Thought experiments as mental modelling. In *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association, Vol. 2*: 291-301. Michigan: Philosophy of Science Association.
- Norton, John D. 2004. Why thought experiments do not transcend empiricism. In *Contemporary Debates in Philosophy of Science*, ed. by Christopher Hitchcock, 44-66. Oxford: Blackwell.
- Parfit, Derek. 1991. *Reasons and Persons*. Oxford: Clarendon Press.
- Peijnenburg, Jeanne and David Atkinson. 2003. When are thought experiments poor ones? *Journal for General Philosophy of Science* 34: 305-322.
- Peijnenburg, J. and Atkinson, D. (2007). On poor and not so poor thought experiments. A reply to Daniel Cohnitz. *Journal for General Philosophy of Science* 38: 159-161.
- Polya, George. 1948. *How to Solve It*. Princeton: Princeton University Press.
- Polya, George. 1954. *Patterns of Plausible Inference*. London: Oxford University Press.
- Rescher, Nicholas. 1976. *Plausible Reasoning*. Assen and Amsterdam: Van Gorcum.
- Rescher, Nicholas. 1987. How serious a fallacy is inconsistency? *Argumentation* 1: 303-316.
- Rescher, Nicholas. 2005. *'What if?': Thought Experimentation in Philosophy*. New Brunswick, N.J.: Transaction Publishers.
- Scott, Sam. 2000. Dueling theories: Thought experiments in cognitive science. In *Proceedings of the Twentieth-Second Annual Conference of the Cognitive Science Society, August 13-15, 2000*, ed. by Lila R. Gleitman and Aravind K. Joshi, 895-900. Philadelphia, PA: Institute for Research in Cognitive Science University of Pennsylvania.
- Searle, John R. 1969. *Speech acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Searle, John R., and Vanderveken, Daniel. 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.
- Sorensen, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.
- Sperber, Dan and Deirdre Wilson. 1995. *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.

- Vanderveken, Daniel. 1990. *Meaning and Speech Acts: Vol. I. Principles of Language Use*. Cambridge: Cambridge University Press.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford: Blackwell.
- Walton, Douglas N., Chris Reeds and Fabrizio Macagno. 2008. *Argumentation Schemes*. New York: Cambridge University Press.

Appendix

"**Suppose** that I am an American soldier in the Second World War and that I am captured by Italian Troops. And **suppose** also that I wish to get these troops to believe that I am a German soldier in order to get them to release me. What I would like to do is to tell them in German or Italian that I am a German soldier. But let us **suppose** I don't know enough German or Italian to do that. So I, as it were, attempt to put on a show of telling them that I am a German soldier by reciting those few bits of German I know, trusting that they don't know enough German to see through my plan. Let us **suppose** I know only one line of German which I remember from a poem I had to memorise in a high school German course. Therefore, I, a captured American, address my Italian captors with the following sentence: *Kennst du das Land, wo die Zitronen blühen?* Now, let us describe the situation in Gricean terms. I intend to produce a certain effect in them, namely, the effect of believing that I am a German soldier, and I intend to produce this effect by means of their recognition of my intention. I intend that they should think that what I am trying to tell them is that I am a German soldier. But does it follow from this account that when I say, *Kennst du das Land...etc.*, what I mean is 'I am a German soldier'? Not only does it not follow, but in this case I find myself disinclined to say that when I utter the German sentence what I mean is 'I am a German soldier', or even 'Ich bin ein deutscher Soldat', because what the words mean and what I remember that they mean is 'Knowest thou the land where the lemon trees bloom?' Of course, I **want** my captors to be deceived into thinking that what I mean is: 'I am a German soldier', but part of what is involved in that is getting them to think that that is what the words I utter mean in German. In the *Philosophical Investigations*, Wittgenstein (discussing a different issue) writes 'Say 'it's cold here' and *mean* 'it's warm here''. The reason we are unable to do this without further stage setting is that what we can mean is at least sometimes a function of what we are saying. Meaning is more than a matter of intention, it is also at least sometimes a matter of convention. One might say that on Grice's account it would seem that any sentence can be uttered with any meaning whatever, given that the circumstances make possible the appropriate intentions. [...] Grice's account can be amended to deal with counter-examples of this kind. We have here a case where I intend to produce a certain effect by means of getting the hearer's recognition of my intention to produce that effect, but the device I use to produce this effect is one which is conventionally, by the rules governing the use of that device, used as a means of producing quite different illocutionary effects, and the stage setting or conditions which would permit us to say one thing and mean something totally unrelated are not present. We **must** [...] reformulate the Gricean account of meaning in such a way as to make it clear that one's meaning something when one utters a sentence is more than just randomly related to what the sentence means in the language one is speaking. In our analysis of illocutionary acts, we **must** capture both the intentional and the conventional aspects and especially the relationship between them." (Searle 1969: 44-45; italics as in the original, bold emphasis added)