The puzzle of thought experiments in conceptual metaphor research

András Kertész

MTA-DE Research Group for Theoretical Linguistics

Abstract
How can thought experiments lead to new empirical knowledge if they do not make use of empirical information? This puzzle has been widely discussed in the philosophy of science. It arises in conceptual metaphor research as well and is especially important for the clarification of its empirical foundations. The aim of the paper is to suggest a possible solution to the puzzle of thought experiments in conceptual metaphor research. The solution rests on the application of a novel metatheoretical framework that conceives of linguistic theorizing as a process of plausible argumentation. The central idea of the solution is that through the dynamic feedback mechanism of the cyclic, prismatic and retrospective re-evaluation of information, thought experiments in conceptual metaphor research may indirectly supply the process of plausible argumentation with empirical knowledge.

Keywords: thought experiment, real experiment, argumentation, plausible inference, metaphor

1.   Introduction

1.1.   The problem

Over approximately the last two decades, thought experiments have become an attractive topic in the philosophy of science.[1] The reason why they are considered challenging is that they raise an epistemological problem called "the paradox of thought experiments" (Kühne 2005: 25), or "the puzzle of thought experiments" (ibid.), or "Kuhn's paradox" (Horovitz & Massey 1991: 1), or "the epistemological problem of thought experiments" (Norton 2004: 44) etc. According to Horovitz and Massey, the paradox was realized by Kuhn who asked: "*How*, […] relying upon familiar data, can thought experiment lead us to new knowledge or to new understanding of nature?" (Kuhn 1977: 241). In Horovitz and Massey's formulation, Kuhn's paradox says that "[…] thought experiments often have *novel empirical import* even though they are conducted entirely inside one's *head* […]" (Horovitz & Massey 1991: 1; emphasis added). That thought experiments yield new empirical knowledge seems to be in conflict with the tenet that gaining empirical knowledge is not possible without empirical data. So, according to the authors mentioned, there is a tension between the circumstance that thought experiments seem to lead to *new empirical knowledge* of the world and the assumption that they achieve this *without making use of experience*. Needless to say, this puzzle concerns the very foundations of scientific research and is, therefore, of utmost importance for the evaluation of empirical inquiry.[2]

To give the reader a first idea of what is meant by a successful thought experiment in the natural sciences, let us mention the most widely discussed example, namely, Galileo's (1974 [1638]: 66f.) story which shows that all falling bodies irrespective of their weight fall at the same speed (for summaries see e.g. Brown 1991: 1, Kühne 2005: 29, Brown and Fehige 2011: 12-13.). The view that heavier bodies fall faster than lighter bodies is attributed to Aristotle. In his famous thought experiment, Galileo challenges this hypothesis by assuming that a heavier canon ball is attached to a lighter musket ball. Then, on the one hand, this pair of objects should fall faster than the canon ball, because they are together heavier than the canon ball alone. On the other hand, they should fall more slowly than the canon ball alone, because the lighter musket ball is expected to slow up the canon ball. Consequently, one obtains the result that the canon ball falls both faster and more slowly than the combination of the canon ball and the musket ball. However, this is an obvious contradiction. The contradiction can be avoided if one assumes that all bodies fall with the same speed. Accordingly, the Aristotelian law of falling bodies has been refuted and replaced by a new natural law. Thus, the thought experiment is provoking because it led to two empirical findings: namely, that the law attributed to Aristotle is false, and that a new natural law of falling bodies is correct. In addition, these new findings have been achieved without considering new empirical data.

Although the puzzle of thought experiments has been raised with respect to the natural sciences, it affects several other fields of research, too. It arises in *conceptual metaphor research* as well. During the past three decades, conceptual metaphor research has become one of the most popular fields in linguistics. Based on Lakoff and Johnson (1980), its leading idea is that the way in which one conceptualizes the world is basically metaphorical. Thereby, it is 'conceptual metaphors' that are decisive. Conceptual metaphors map a concrete concept

---

[1] For overviews of the literature on thought experiments see Brown and Fehige (2011), Sorensen (1992), Kühne (2005), Buzzoni (2008), Cohnitz (2006), Brown (1991), Gendler (2000), Moue et al. (2006).
[2] Throughout the paper, we use the term 'empirical' pre-explicatively.

(called 'source domain') onto an abstract one (the 'target domain'), and thus the latter is understood in terms of the former. For example, the conceptual metaphor ARGUMENT IS WAR projects the concrete conceptual domain of WAR on the abstract conceptual domain of ARGUMENT.[3] Then, we understand arguments in terms of war. This conceptual metaphor is manifested in such linguistic expressions as *she defended her position*; *she attacked my standpoint* etc.

In the classical contributions to conceptual metaphor research, *thought* experiments played a constructive role and *real* experiments were not conducted at all. Since these contributions claim to be empirical while they are built on thought experiments,[4] they are directly affected by the puzzle. Against this background, we raise the puzzle of thought experiments in conceptual metaphor research as follows:

(P)   *How* can thought experiments in conceptual metaphor research yield new empirical knowledge of metaphors, if they do not make use of new empirical information that goes beyond what has already been known?

The relevance of (P) is that its solution is *indispensable for the clarification of the empirical foundations* of conceptual metaphor research.

1.2.   *The structure of the paper*

In order to find a possible solution to (P), we will proceed as follows:

First of all, a framework will be needed that is capable of providing a possible solution to (P). In Section 2, we will argue for *a novel metatheoretical framework* that we expect to yield our solution to (P) – namely, the *p-model* of Kertész and Rákosi (2012).

In order to exemplify the workability of the argumentation theoretical framework, Section 3 will be devoted to *a case study* on one of Lakoff and Johnson's (1980) classical thought experiments. Thereby, we have chosen the very thought experiment with which the book starts and which obviously has the function to convince the reader of the tenability of the theses defended in the book. The thought experiment includes the conceptual metaphor ARGUMENT IS WAR that we have already mentioned above in order to illustrate the leading idea of conceptual metaphor research.

In Section 4, we will summarize the substantial properties of the thought experiment at issue that the case study has revealed. In opposition to current philosophical approaches to thought experiments, we will conclude that they must not be restricted to the 'story' they tell, but they comprise a *complex phase within a process of plausible argumentation* that consists of cycles of the retrospective re-evaluation of given assumptions from new perspectives. Second, the application of the p-model will also show that the main function of thought experiments in conceptual metaphor research is to increase the plausibility of certain hypotheses and to decrease that of their rivals, to test the plausibility of the alternative hypotheses, and to serve as a special kind of evidence for or against them.

In Section 5, we will conduct a *second case study*. The motivation for doing so is that the philosophical literature tries to capture the nature of thought experiments by both comparing them to and separating them from real experiments (see Buzzoni 2008). Therefore, it is reasonable to assume that the way real experiments in conceptual metaphor research work

---

[3] In conceptual metaphor research, concepts are represented by capital letters or small capitals.

[4] "Empirical study is necessary." (Lakoff and Johnson 1999: 5)

will contribute to finding a solution to our central problem (P). In the second case study, we will analyze real experiments reported on in Gibbs et al. (2004) that are related to the thought experiment discussed in Section 3. Gibbs et al. (2004) includes an experimental report on two real experiments the aim of which was to test the tenet of Lakoff and Johnson's conceptual metaphor theory according to which the source domains of conceptual metaphors are rooted in bodily experiences and reflected in the abstract target domains. As an example, in the real experiments the authors examine the conceptual metaphor DESIRE IS HUNGER and argue for the finding that the more prominent parts of the experimentees' hunger experiences are invariantly mapped onto their concepts of desire.

Based on our analysis of the experimental report, in Section 6 we will gain a series of insights into the relationship between thought experiments and real experiments in conceptual metaphor research. In particular, we will conclude that there is a cyclic relationship between thought experiments and real experiments. First, the results of thought experiments may be part of the plausible argumentation process that real experiments consist of. Second, through the cycles of the retrospective re-evaluation of information, the result of real experiments may retrospectively modify the plausibility of the result of the thought experiment that has been included in the starting context of the real experiments. Third, real experiments may include thought experiments.

Finally, in Section 7, we will derive our answer to the *how*-question as asked in (P) from the above findings. The central idea of our solution is that through the dynamic feedback mechanism of the cyclic, prismatic and retrospective re-evaluation of information, thought experiments in conceptual metaphor research may indirectly supply *the process of plausible argumentation* with empirical knowledge.

1.3.   *On current research on thought experiments and the novelty of the present approach*

There seems to be agreement on the fact that striving for a generally acceptable definition of the notion of 'thought experiment' is a hopeless endeavor (see e.g. Brown 1991: 1, Kühne 2005: 23-24, Peijnenburg and Atkinson 2003: 306). What most authors seem to agree upon is merely that (a) thought experiments focus on some *imaginary situation*, (b) are *hypothetical* in the sense that they give an answer to a 'What-if' question, and (c) are *exemplified* by a number of cases mainly from the field of physics. We will not attempt to give a general definition of the notion of 'thought experiment', either. Rather, in order to provide a first insight into the novelty of the main ideas that the present paper is intended to argue for, here we will refer briefly to current solutions to the puzzle of thought experiments and then indicate how our approach differs from them.

The two best-known attempts are those by Brown and Norton, and the debate between them is one of the highlights of the philosophical literature on thought experiments. Norton's solution belongs to a group of approaches that consider thought experiments to be 'inferences' or, to use an alternative term, 'arguments' (Norton 2004, Häggquist 1996, Cohnitz 2006 etc.). Norton's solution to the puzzle of thought experiments maintains that "[…] in so far as a thought experiment provides novel information about the world, that information was introduced as *experientially based premises* in the arguments" (Norton 2004: 63; emphasis added). Accordingly, thought experiments are not mystical and do not undermine empiricism. Brown considers only what he calls 'Platonic thought experiment'. A Platonic thought experiment "is *a priori* in that it is not based on new empirical evidence nor is it merely logically derived from old data; and it is an advance in that the resulting theory is better than the predecessor theory." (Brown 1991: 77).

Another group of approaches (see e.g. Gooding 1990 and McAllister 1996) assumes that there is no significant difference between thought experiments and real experiments. The main argument for this claim is that the essential feature of both real and thought experiments is the same, namely, that they manipulate something – the only minor difference being that while real experiments manipulate objects, thought experiments manipulate thoughts.

A fourth kind of approaches claims that thought experiments construct a mental model of the world (Nercessian 1993, Miščević 1992, Cooper 2005). This assumption provides a solution to the puzzle according to which "thought experiments can show us whether or not a situation is possible. In doing this they can indirectly teach us about the actual world. Discovering that a situation is impossible shows us how the world cannot be. Similarly, discovering that a situation is necessary shows us how the world must be" (Cooper 2005: 339).

In linguistics, the only relatively systematic attempt to raise the issue of thought experiments was made in Thomason (1991). She considers many examples of thought experiments taken from different linguistic approaches. She distinguishes between two kinds of linguistic thought experiments. The first is what she calls *stage-stetting thought experiments*. A stage setting thought experiment works as a first step in an argumentation in order to clarify the theoretical issue and to get the audience's agreement before the research itself has been carried out. Then it is followed by a second step demonstrating that there is a real-world situation similar to the result of the thought experiment.[5]

The second kind is *introspective thought experiments* about the correctness of linguistic structures. They are carried out by informants or the linguist herself. They function as "actual tests of hypotheses about language structure" (Thomason 1991: 253).[6]

After this sketchy overview of current approaches to solve the puzzle of thought experiments, let us briefly indicate how the specific solution to be inferred from the investigations reported on in the present paper differs from them. As a result of our line of reasoning to be carried out in the sequel, we will obtain the following solution to the problem (P):

(SP)   (a)   Thought experiments in conceptual metaphor research are to be conceived of as complex phases in the dynamic process of plausible argumentation.

---

[5] As an example, Thomason mentions one of the problems of language acquisition: "What sorts of inherited mental structures must be posited to account for the fact that human infants learn their first language with spectacular success in the first few years of their lives?" (Thomason 1991: 248). To answer this question, the linguist conducts a thought experiment by hypothetically assuming that an infant born in a completely homogenous monolingual environment was removed from this environment into another one in which his/her adoptive parents speak a language completely unrelated to the native language of the child's biological parents. The linguist's conclusion is that the child will learn the language of his/her new environment in the same way and with the same speed as if he/she were still in his/her original linguistic environment. As a next step, the linguist may find similar situations in everyday life and try to compare them to this imaginary situation thus checking the conclusion of the thought experiment. In this way, the thought experiment has the task to set the stage for further argumentation steps and is to be continued by the examination of real-world situations.

[6] Examples of introspective thought experiments include grammaticality judgments in generative syntax or the identification of the meaning of a given expression in cognitive semantics etc.

(b)   In accordance with this, new empirical information is obtained from sources which are outside the thought experiment, but which have been put into the context by the overall cyclic, prismatic and retrospective process of plausible argumentation.

(c)   Thought experiments may transmit empirical information from the context to the particular cycles of the process of plausible argumentation.

(d)   Accordingly, a thought experiment cannot provide new empirical information itself, but the novel insights it yields consist in its contribution to the cyclic, prismatic and retrospective re-evaluation of empirical information.

(e)   Thereby, the already known knowledge put into this process is not identical with its output, because in the course of the cyclic, prismatic and retrospective re-evaluation, the plausibility value of the statements at issue changes dynamically.

The novelty of (SP) is that, first and most importantly, its central idea differs from all the approaches mentioned, because none of them models thought experiments as phases within a dynamic process of plausible argumentation. In particular, whereas Norton's approach is restricted to single inferences, our solution focuses on the much broader notion of 'plausible argumentation' that consists of a series of interconnected inferences and involves the continuous re-evaluation of information. In addition, the claim that thought experiments cannot provide new empirical information themselves, is anything but trivial, since it clearly differs both from the above quotation from Norton (2004) and Horovitz and Massey's formulation we have also quoted in the first paragraph of Subsection 1.1. Similarly to Gooding's (1990) and McAllister's (1996) approach, the p-model accepts the analogy between thought experiments and real experiments but, unlike the latter, it captures this analogy within the overall process of plausible argumentation. The p-model does not make use of the notion of 'mental model' in the sense of Cooper (2005) or Nercessian (1993). Unlike Thomason (1991), we will see that introspective and stage setting thought experiments in conceptual metaphor research are closely related. Finally, Brown's solution is fully incompatible with (SP).

Second, according to (SP), the main function of thought experiments in conceptual metaphor research is to *increase the plausibility* of certain hypotheses and to *decrease* that of their rivals. None of the approaches mentioned makes use of the notion of plausibility.

Third, thought experiments in conceptual metaphor research serve as a special kind of evidence – that we will call 'relative evidence' – for a certain hypothesis and against its rival.

Fourth, in the philosophical literature the relationship between thought experiments and real experiments is anything but clarified (for an overview of this issue see Buzzoni 2008). In accordance with (SP), this relationship is at least threefold: thought experiments may be the input to real experiments, real experiments may retrospectively modify the plausibility of the outcome of thought experiments, and thought experiments may be components of real experiments.

After having overviewed the research question as well as the structure and the main tenets of the paper in this introductory section, let us turn to the research itself now. To start with, in the next section we will have to introduce the research framework.

2. The metatheoretical framework

We hypothesize that the application of Kertész and Rákosi's (2012) *p-model of plausible argumentation* may provide an appropriate metatheoretical framework for the solution of (P). Before introducing some of its central notions, two remarks should be made.

First, the p-model is an argumentation theoretical approach to linguistic inquiry focusing on the data/evidence problem widely discussed in linguistics.[7] It was developed independently of thought experiments.[8] Accordingly, it does not claim general validity at the outset and does not strive for a general definition of 'argumentation'; rather, it may be the task of further case studies to examine whether, and if so, how and to what extent the p-model can be applied to other fields (i.e. everyday argumentation or disciplines other than linguistics).

Second, the main advantage of its application to thought experiments in conceptual metaphor research would be that it could capture them with the same means that are also applicable to some other relevant aspects of linguistic inquiry. Accordingly, it would model thought experiments in conceptual metaphor research as constitutive components of linguistic inquiry, on a par with other components.

Below we sketch the general features of the p-model in a simplified, informal manner. The notions to be introduced constitute a hierarchy. Those introduced later presuppose the ones introduced earlier. Basically, the p-model centers on the notion of 'plausible argumentation' (see point (ix) below); all other notions to be introduced are building blocks from which this notion has been constructed. For the precise definitions see Kertész and Rákosi (2012).

(i) Plausible statements. In scientific inquiry, it is very often the case that the statements one deals with are not truths with certainty, but they are merely *plausible* to some extent. A plausible statement consists of a proposition and a plausibility value. The plausibility value assigned to a statement depends on the *reliability of the sources* that it stems from. For example, the statement "*We have found [...] that metaphor is pervasive in everyday life, not just in language but in thought and action*" (Lakoff and Johnson 1980: 3) is merely plausible. Its plausibility degree depends on the reliability of the source in comparison to its alternatives: in this case, the source is what the authors refer to by the expression "*we have found that*", i.e. the investigations they carried out. This source assigns a higher plausibility to the claim that metaphor is pervasive in everyday life etc. than to the rival claim that metaphor is a device of poetic imagination (see (2) vs. (3) in Subsection 3.1).

(ii) Sources. We distinguish between two kinds of sources. A source is *direct* with respect to a statement if the statement is assigned a plausibility value based on the reliability of the given source. In this sense, corpora, theories, conjectures, the intuition of native speakers, experiments, thought experiments, fieldwork, historical documents, dictionaries and videotapes are direct sources. If the plausibility value of the given statement depends on an inference, then we speak of *an indirect source*. For example, if in the above statement the expression "*we have found that*" referred to the authors' intuition, or a scientific paper, or statistical data, then the source of its plausibility would be direct. However, if the same expression referred to an inference the conclusion of which is the statement quoted, then the source of the plausibility of this statement would be, in the sense mentioned, indirect. We introduce the notational convention according to which plausible statements will be set within

---

'|' and sources will be indicated by subscripts. Thus, '$|p|_S$' should be read as 'the statement p is plausible on the basis of the source S'.

(iii) *Plausibility values*. The plausibility value of statement p on the basis of the source S is such that:

$|p|_S = 1$, if p is true with certainty on the basis of S;
$0 < |p|_S < 1$, if p is plausible on the basis of S;
$0 < |\sim p|_S < 1$, if p is implausible on the basis of S;
$|p|_S = 0$, 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:[9]

$0 <$ |Metaphor is pervasive in everyday life, not just in language but in thought and action.$|_S < 1$

Here S stands for the investigation carried out in Lakoff and Johnson (1980) as a direct source. The plausibility value of statements is not absolute, but gradual and comparative. It may be the case that a particular statement is plausible with respect to a given source and less plausible relative to another; or, the same source may assign a higher plausibility value to a given statement than to its rival. Kertész and Rákosi (2012) use numerical scales of plausibility values that, for the sake of simplicity, we do not introduce in the present paper.

(iv) *Plausible inferences*. The basic idea underlying plausible inferences can be best understood if we compare them to deductive and demonstrative ones. *Deductive* inferences *hypothetically* assume the truth of their premises: if the premises are true, then, by virtue of the logical consequence relation between the premises and the conclusion, the conclusion is also true. However, in scientific inquiry one often wants to arrive at conclusions that are actually true. Such conclusions are obtained by *demonstrative* inferences. Demonstrative inferences go beyond deductive ones insofar as, besides retaining the logical consequence relation between the premises and the conclusion and thus being deductively valid in the same sense as deductive inferences, their premises are certainly true and therefore, their conclusion is true with certainty, too. Nevertheless, in scientific practice such a desideratum is in most cases not available at all, because either there is no logical consequence relation between the premises and the conclusion, or at least one of the premises is merely a plausible statement, or both. Then the conclusion is not true with certainty, but plausible, too. This kind of inference is called *plausible inference*. In plausible inferences the connection between the premises and the conclusion involves *a semantic relation* such as causality, analogy, necessary or sufficient condition, part-whole relation, whole-part relation etc. For example: *Grammar is similar to computer programs; computer programs are modularly organized;* therefore: *Grammar is modularly organized.*

(v) *Three types of plausible inference*. As regards the structural relation between the premises and the conclusion, there are three basic types of plausible inference. The first is characterized by the fact that at least one of the premises is not true but only plausible, and accordingly, the conclusion is merely plausible, too. For example, in the following inference both premises are merely plausible, and, accordingly, the conclusion is not true, either: *If the sky is dark, then it will be raining*; *The sky is dark*; therefore: *It will be raining*. In the second type, the premises are true, but there is no logical consequence relation between them and the

---

[9] In order to make the notions to be introduced more easily understandable, in the illustrative examples of the next points we will not use this notation. However, its use will be important in the case studies.

conclusion. Therefore, *latent background assumptions* have to be added to the premises in order to establish the logical consequence relation between the premises and the conclusion. For example: *John's body is big*; *John's head is part of his body*; therefore: *John's head is big*. Here the missing background assumption is: *If the whole has property P, then the part also has property P*. Inferences including latent background assumptions besides the premises are called *enthymematic*.[10] The third type is the combination of the first two in that it is enthymematic and includes plausible statements as premises. For example: *Knowledge of language is organized in a modular way*; *Knowledge of language is part of human cognition*; therefore: *Human cognition is organized in a modular way*. Here both premises are merely plausible and the inference is enthymematic.

(vi) The p-context. The workability of plausible inferences depends on many factors, among others on all pieces of information that may be relevant for the plausibility value of the premises and the latent background assumptions. The totality of such information is called the *p-context*. The prefix 'p' serves to distinguish this kind of context from the notion of 'context' without a prefix as used in linguistics. Plausible inferences are p-context-dependent. Basically, the p-context contains three kinds of information: first, the methodological rules pertaining to the permissible inference types, the methods of the treatment of contradictions, the criteria for the judgment of the reliability of sources etc.; second, the direct and indirect sources considered to be reliable on the basis of the accepted methodological rules; and third, plausible statements along with their relevant characteristics at one's disposal according to the methodological rules and the sources. From these characteristics it follows that the p-context differs in at least two respects from the notion of context as used e.g. in pragmatics: only statements that are relevant for the plausibility of other statements belong to the p-context, and the p-context may include several instances of a given statement whose plausibility value is supported by different sources. For example, in the experiments we will analyze in the second case study in Section 5, the p-context contains the following methodological principle: *Empirical research makes use of real experiments*. Then, this methodological principle serves as a latent background assumption in a plausible inference (here and in what follows latent background assumptions will be put within '[' and ']'): [*Empirical research makes use of real experiments.*]; *Empirical research is needed to establish connections between embodiment and metaphor in thought and language*; therefore: *Real experiments are needed to establish connections between embodiment and metaphor in thought and language*. The context may change dynamically – as we will see in the case studies.

(vii) The informational over- and underdetermination of the p-context. Since, as mentioned, p-contexts may be very rich in that they include different kinds of information influencing the plausibility of statements, it may happen that a statement is made plausible by a particular source while its negation is supported by another source. Such p-contexts are informationally *over*determined and the set of the plausible statements in the p-context is *inconsistent*. However, the opposite case may also occur: namely, informational *under*determination. An informationally underdetermined p-context is incomplete in the sense that it includes at least one statement that is neither plausible nor implausible relative to the

---

[10] See e.g. Rapp (2010) for the history of this notion. See Rescher (1976: 60ff.), Polya (1948: 223) for enthymematic plausible inferences. For an overview of the problems they raise and the argumentation schemes in which they participate see Walton et al. (2008). Nevertheless, the identification of the latent background assumptions is not without problems: "One 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 […]" (Walton et al. 2008: 189).

sources available. A p-context may be both informationally under- and overdetermined (with respect to different statements).

(viii) P-problems, their solution and their resolution. Informationally over- and/or underdetermined p-contexts are called *p-problematic*, and instances of informational over- and underdetermination are *p-problems*. P-contexts are p-problematic if they affect the plausibility of a statement that is relevant to the inquiry at issue. A *solution of a p-problem* is a p-context in which the statement in question is either plausible or implausible with respect to all sources supporting it. However, p-problems may have several solutions supported by different sources. The *resolution of a p-problem* is a solution of the given p-problem which, when compared with other solutions, is the most plausible, relative to the information at one's disposal.

(ix) Plausible argumentation. In order to reach the solution/resolution of a given p-problem, a heuristic tool is needed. We call this tool *plausible argumentation*. We define 'plausible argumentation' as the gradual transformation of a p-problematic p-context into a less p-problematic one by the use of a series of plausible inferences. It involves a *dynamic process* consisting of a sequence of plausible inferences in the course of which a p-problematic p-context is continuously *re-evaluated* by the elaboration of possible solutions to the p-problems at issue, the consideration of rival solutions and the comparison of the latter. After all the available solutions have been evaluated, the process of plausible argumentation culminates in deciding which of the rivals to accept as the resolution of the given p-problem. Nevertheless, even the resolution of the p-problem may be merely provisional and may be re-evaluated in later phases of the process.

The re-evaluation of a p-problematic p-context only rarely leads immediately to an unproblematic one and may raise new p-problems. Therefore, the consideration of new alternatives and the revision of previous decisions are needed. Previous decisions are not static. One continuously returns to previous p-problems, re-evaluating earlier decisions concerning the acceptance or rejection of statements, the reliability of the sources, the plausibility values of the statements, the workability of methodological norms, previously drawn inferences etc. To put it simply: during the argumentation process the pieces of information one deals with are *retrospectively re-evaluated* (cf. also Rescher 1976, 1987).

The retrospective re-evaluation of information is carried out in sequences of *cyclic* returns. Moreover, the cycles continuously change the perspective from which the information in the p-context is evaluated. In this sense, the cycles of the retrospective re-evaluation are *prismatic* – they are viewed through the 'prisms' of later phases of the argumentation process (cf. also Rescher 1987).[11]

(x) The phases of the plausible argumentation process. An argumentation process begins with the *starting p-context* that is p-problematic in the above sense and includes the initial p-problem, the initial sources, hypotheses and background assumptions. Then, one or several argumentation cycles may follow in which the p-context is *elaborated* for example in that it is extended by new sources, new statements, new methods; or it is modified by the deletion of the sources which no longer seem to be reliable. After the last cycle, one reaches the *final p-context* defined by the (provisional) solution or resolution of the p-problem raised in the starting p-context. (As an example, see Section 3.)

In the next section, we will carry out a simple case study on one of Lakoff and Johnson's thought experiments. Its aim is twofold. First, it will exemplify how the p-model can be applied to the analysis of thought experiments in conceptual metaphor research.

---

[11] Rescher uses the term 'retrovalidate'. We do not accept this term, because in most cases the rerospective process does not lead to truths, but merely modifies the plausibility value of statements. The term 're-evaluate' captures this characteristric of the process more precisely.

Second, the case study will prepare conclusions concerning the structural and functional properties of thought experiments in conceptual metaphor research.

3. Case study 1: The argumentation structure of a thought experiment in conceptual metaphor research

3.1. The starting p-context

In the starting p-context, Lakoff and Johnson raise the initial p-problem right in the first sentence of the Preface:

(1)   "This book grew out of a concern, on both our parts, with *how people understand their language and their experience*". (Lakoff and Johnson 1980: ix; emphasis added)

Then, they put forward the two alternative hypotheses as possible solutions to (1). The first is the view that, according to the authors, most people share:

(2)   "Metaphor is for *most people* a device of the poetic imagination and the rhetorical flourish – a matter of extraordinary rather than ordinary language. Moreover, metaphor is typically viewed as characteristic of language alone, a matter of words rather than thought or action." (Lakoff and Johnson 1980: 3; emphasis added)

The source of (2) is part of the starting p-context (in the sense of (vi) and (x) in Section 2) of the authors' argumentation and is identified with "*most people*". It is a direct source. However, they contrast (2) with an alternative hypothesis the plausibility of which is at this stage of the argumentation very low:

(3)   "*We have found*, on the contrary, that metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature." (Lakoff and Johnson 1980: 3; emphasis added).

Here the source of plausibility is "*we have found*", but it is unclear whether it is direct or indirect in the sense of (iii) in Section 2. At this stage of the argumentation process one cannot know what exactly this source is: is it corpora, real experiments, thought experiments, intuition, authority, inferences, etc.?
    Obviously, the starting p-context including (2) and (3) is p-problematic in the sense of (vii) in Section 2 because it is informationally overdetermined. Namely, the authors assume that they are inconsistent.[12] Thus, the authors' aim is to increase the plausibility value of their main thesis (3) and to make its rival as quoted in (2) implausible.

---

[12] Please notice that it is not the case that (2) and (3) are complementary claims, (2) about a common opinion and (3) about research findings. First, „most people" may include scholars as well. Second, point (g) of the quotation in (4) below clearly witnesses that the authors treat (2) and (3) as rival hypotheses that cannot be accepted simultaneously. Third, in point (vii) in Section 2 we said that if a statement is made plausible by a particular source while its negation is supported by another source, then the p-context is informationally overdetermined i.e. inconsistent. Accordingly, (2) and (3) are *rival hypotheses*.

3.2.   The elaboration of the p-context

Thought experiments are not directly observable. One can access them only via their verbal presentation that we will call 'thought experimental report' in analogy to the term 'experimental report' used with respect to real experiments. In order to argue against (2) and for (3), that is, to transform the p-problematic starting p-context into a less p-problematic one, Lakoff and Johnson *extend* the starting p-context (in the sense of (x) in Section 2) by introducing the following *thought experimental report*:[13]

(4)   (a)   "Try to imagine a culture where arguments are not viewed in terms of war, where no one wins or loses, where there is no sense of attacking or defending, gaining or losing ground. Imagine a culture where an argument is viewed as a dance, the participants are seen as performers, and the goal is to perform in a balanced and aesthetically pleasing way.

(b)   In such a culture, people would view arguments differently, experience them differently, carry them out differently, and talk about them differently.

(c)   But *we* would probably not view them as arguing at all: they would simply be doing something different. It would seem strange even to call what they were doing 'arguing'.

(d)   Perhaps the most neutral way of describing this difference between their culture and ours would be to say that we have a discourse form structured in terms of battle and they have one structured in terms of dance.

(e)   This is an example of what it means for a metaphorical concept, namely, ARGUMENT IS WAR, to structure (at least in part) what we do and how we understand what we are doing when we argue.

(f)   *The essence of metaphor is understanding and experiencing one kind of thing in terms of another*. […] ARGUMENT is partially structured, understood, performed, and talked about in terms of WAR. The concept is metaphorically structured, the activity is metaphorically structured, and, consequently, the language is metaphorically structured.

(g)   The most important claim we have made so far is that metaphor is not just a matter of language, that is, of mere words. We shall argue that, on the contrary, human thought processes are largely metaphorical. This is what we mean when we say that the human conceptual system is metaphorically structured and defined." (Lakoff and Johnson 1980: 4-5; emphasis as in the original.)

As it was shown in Kertész and Rákosi (2009), the crucial point in (4)(a) and (b) is that the following *latent background* assumption (in the sense of (v) in Section 2) is not formulated explicitly in the thought experiment, although it is part of the p-context:

(5)   There are systematic connections among thinking, experience, action and language.

In what follows, we will take up the seven points (4)(a)-(g) individually in order to reconstruct the argumentation structure of the thought experiment quoted. Let us begin with the reconstruction of the inference in (4)(a) and (b). In accordance with points (ii) and (iii) in

---

[13] We will divide the quotations into smaller units in order to make it easier to refer to parts of the quotation to be analyzed. The units do not necessarily correspond to paragraphs in the original text.

Section 2, here and in what follows, plausible statements are set within '|'. Outside the latter there are the plausibility values assigned to the statement based on a given source. In the subscripts, *S(4)(a)* stands for the imaginary situation in *(4)(a)* as a direct source which has been introduced into the p-context of Lakoff and Johnson's argumentation, and which supports the plausibility of the statement at issue. *I(6)* stands for the inference in (6) as an indirect source of (6)(d) in the sense of (ii) in Section 2. Latent background assumptions are set within '[' and ']'.

(6)   Premises: [14]
    (a) $0<$|In an imaginary culture argument is not viewed in terms of war, but in terms of dance.|$_{S(4)(a)}<1$
    (b)   [$0<$|There are systematic connections among thinking, experience, action and language.|$_{S(4)(a)}<1$]
    (c)   [$0<$|If we imagine a culture in which argument is not viewed in terms of war, but in terms of dance and there are systematic connections among thinking, experience, action and language, then in this culture people would view arguments differently, experience them differently, carry them out differently, and talk about them differently.|$_{S(4)(a)}<1$]
    Conclusion:
    (d) $0<$|In this culture people would view arguments differently, experience them differently, carry them out differently, and talk about them differently.|$_{I(6)}<1$

The imaginary situation as reconstructed above is further analyzed in (4)(c) where it is compared to our culture:

(7)   Premises:
    (a) $0<$|People in this culture view arguments differently, experience them differently, carry them out differently, and talk about them differently.|$_{S(4)(a)}<1$
    (b)   [$0<$|There are systematic connections among thinking, experience, action and language.|$_{S(4)(a)}<1$]
    (c)   [$0<$|If people in this culture viewed arguments differently, experienced them differently, carried them out differently, and talked about them differently, and there are systematic connections among thinking, experience, action and language, then we probably would not view them as arguing at all, they would simply be doing something different which we would not call 'arguing'.|$_{S(4)(a)}<1$]
    Conclusion:
    (d) $0<$|We probably would not view them as arguing at all, they would simply be doing something different which we would not call 'arguing'.|$_{I(7)}<1$

As regards (4)(d), in the structure of the inference there are latent background assumptions related to the notion of 'discourse form', a concept that has not been defined explicitly:

(8)   Premises:
    (a)   [$0<$|If one views argument in terms of dance, then one has a discourse form structured in terms of dance.|$_{S(4)(a)}<1$]
    (b) $0<$|They view argument in terms of dance.|$_{S(4)(a)}<1$

---

[14] The structure of (6) and (7) corresponds to that of plausible *modus ponens*: $0<$|If A, then B|$_S < 1$; $0<|A|_S < 1$; therefore, $0<|B|_I < 1$.

13

Conclusion:
(c) $0<|$They have a discourse form structured in terms of dance.$|_{I(8)}<1$

(9)   Premises:
(a) $[0<|$If one views argument in terms of battle, then one has a discourse form structured in terms of battle.$|_{S(4)(a)}<1]$
(b) $0<|$We view argument in terms of battle.$|_{S(4)(a)}<1$
Conclusion:
(c) $0<|$We have a discourse form structured in terms of battle.$|_{I(9)}<1$

(4)(e) includes a simple plausible *modus ponens* again:

(10)   Premises:
(a) $0<|$If in the imaginary culture people have a discourse form structured in terms of dance and in our culture we have a discourse form structured in terms of war, then a metaphorical concept, namely, ARGUMENT IS WAR, structures (at least in part) what we do and how we understand what we are doing when we argue.$|_{S(4)(a)}<1$
(b) $0<|$In the imaginary culture people have a discourse form structured in terms of dance and in our culture we have a discourse form structured in terms of war.$|_{S(4)(a)}<1$
Conclusion:
(c) $0<|$A metaphorical concept, namely, ARGUMENT IS WAR, structures (at least in part) what we do and how we understand what we are doing when we argue.$|_{I(10)}<1$

(4)(f) involves an inductive generalization.[15] The point is that (11)(b), (c), (d) etc. are latent background assumptions, that is, it is assumed at this point of the plausible argumentation process that further examples will be found and that all of them behave in the same way; then, the conclusion generalizes over these cases. Indeed, the examples anticipated here will be provided by later considerations in Lakoff and Johnson's book. Accordingly, they are part of the p-context in the sense in which we have introduced this notion (see (vi) in Section 2). Nevertheless, the latent background assumptions are of neutral plausibility at this point of the argumentation process because there is no source available against which their plausibility value could be judged.

(11)   Premises:
(a) $0<|$The essence of the metaphorical concept ARGUMENT IS WAR is understanding and experiencing one kind of thing in terms of another.$|_{S(4)(a)}<1$
(b) $[0<|$The essence of the metaphorical concept X is understanding and experiencing one kind of thing in terms of another.$|_{S(4)(a)}=0]$
(c) $[0<|$The essence of the metaphorical concept Y is understanding and experiencing one kind of thing in terms of another.$|_{S(4)(a)}=0]$
(d) $[0<|$For every other metaphorical concept, its essence is understanding and experiencing one kind of thing in terms of another.$|_{S(4)(a)}=0]$
etc.
Conclusion:

---

[15] Inductive inferences are a subtype of plausible inferences.

(d) $0<|$The essence of metaphor is understanding and experiencing one kind of thing in terms of another.$|_{I(11)}<1$

In (4)(g), Lakoff and Johnson's main hypothesis that we introduced in (3) and its rival as quoted in (2) are compared. The comparison witnesses that the authors consider them to be *rival claims*. (4)(g) includes a part-whole inference:[16]

(12)   Premises:
(a) $0<|$ARGUMENT is partially structured, understood, performed, and talked about in terms of WAR. The concept is metaphorically structured, the activity is metaphorically structured, and, consequently, the language is metaphorically structured.$|_{S(4)(a)}<1$
(b) $[0<|$ARGUMENT IS WAR is part of the human conceptual system.$|_{S(4)(a)}<1]$
(c) $[0<|$If the part has property P, then the whole also has property P.$|_{S(4)(a)}<1]$
(d) $[0<|$The property P is the partial structuring, understanding and performing of something$|_{S(4)(a)}<1]$.
        Conclusion:
(e) $0<|$The human conceptual system is metaphorically structured, the activity is metaphorically structured, and, consequently, the language is metaphorically structured.$|_{I(12)}<1$

3.3.   The final p-context

(12)(e) is equivalent to (3) that is Lakoff and Johnson's main hypothesis. Thus, at this stage of the plausible argumentation process, (12)(e) cyclically, prismatically and retrospectively re-evaluates both Lakoff and Johnson's main hypothesis and its rival against which they argue in the sense of (ix) in Section 2. It decreases the plausibility of the rival claim (2) and increases the plausibility of their own claim (3). As a result, a provisional *solution* – but no resolution – of the p-problem they raised (see (1)) has been achieved.

In the course of the book a series of further argumentation cycles are carried out the results of which retrospectively re-evaluate (12)(e) and through this further increase the plausibility of (3). However, here we restrict our attention merely to the thought experiment in (4), and it would be beside the point to go into the subsequent argumentation stages the book consists of.

As the next step of our line of reasoning, in Section 4 we will draw relevant conclusions from the above case study concerning the structural and functional properties of thought experiments in conceptual metaphor research.

4.   *Summary of the findings of case study 1*[17]

---

[16] The structure of a part-whole inference is: $0<|A$ is part of $B|_S<1$; $0<|A$ has the property $X|_S<1$; therefore, $0<|B$ has the property $X|_I<1$.

[17] This solution to (P2) is restricted to the identification of the *genus proximum* of the structure of thought experiments in conceptual metaphor research, while here we cannot go into the question of what the *differentia specifica* of subtypes of thought experiments in this field of research are.

As the above analysis exemplifies, the thought experiment in (4) has the following relevant *structural* properties:

(a) As mentioned in Section 1, most philosophical approaches restrict a thought experiment to the description of an imaginary situation. The application of the p-model to Lakoff and Johnson's thought experiment has shown that, instead of being restricted to an imaginary situation, it constitutes *a complex phase within* a process of plausible argumentation. It begins with a p-problematic starting p-context the elaboration of which includes its extension by an imaginary situation. It (temporarily) terminates with the final p-context in the sense of (x) in Section 2. The final p-context contains a possible solution of the initial p-problem in (1) in that it triggers the decision between the alternative hypotheses.

(b) The solution to the initial p-problem in (1) is obtained by the use of a series of plausible inferences that yield conclusions with different degrees of plausibility.

(c) Some of the plausible inferences are enthymematic. The latent background assumptions inherent in the p-context – for example, (6)(b), (8)(a), (9)(a), (11)(b), (c), (12)(b) etc. – may contribute to the conclusions of the inferences we reconstructed.

(d) All the inferences into which the thought experiment has been decomposed are components of a process of plausible argumentation. Some of the plausible inferences work as prisms through which during later cycles of the argumentation process previously drawn conclusions are retrospectively re-evaluated. There is a continuous retrospective re-evaluation of previous conclusions by feeding back to them cyclically and prismatically in the light of further information.

(e) The role of the p-context is crucial, because the information used as premises and latent background assumptions is included in the p-context.

(f) One of the components of the plausible argumentation process is the comparison of the imaginary situation with one's own personal experiences of similar real situations. In particular, Thomason (1991: 248) mentions that a real-world situation should be similar to the result of the thought experiment (see also Sorensen 1992: 89 and Cohnitz 2006: 81 on this issue). We reconstruct this comparison as the following analogical inference:[18]

(13)   Premises:
(a)   [$0<|$The imaginary situation as described in (4)(a) is similar to real situations one has already experienced.$|_{SPE}<1$]
(b) $0<|$In the imaginary situation as described in (4)(a), the essence of metaphor is understanding and experiencing one kind of thing in terms of another.$|_{S(4)(a)}<1$
Conclusion:
(c)   [$0<|$In the real situations, one has experienced the essence of metaphor is understanding and experiencing one kind of thing in terms of another.$|_{I(13)}<1$]

(13)(a) integrates previous experience into the process of plausible argumentation from the p-context, and (13)(c) retrospectively re-evaluates this previous experience through the prism of the imaginary situation referred to in (13)(a). *SPE* stands for the personal experience at issue as a direct source.

The structural properties of the thought experiment discussed are closely connected to its *functions*, which can be summarized as follows:

(g) The first function of (4) is stage setting. The thought experiment in (4) works exactly as Thomason (1991: 247-248) characterized stage setting thought experiments. It clarifies the theoretical issue in that it explains what the authors mean by conceptual

---

[18] The structure of analogical inferences is: $0<|A$ is similar to $B|_S <1$; $0<|A$ has the property $X|_S <1$; therefore, $0< |B$ has the property $X|_I <1$.

metaphor. It serves to convince the reader of the book's central hypotheses before the detailed analyses of metaphorical expressions and metaphorical concepts have been carried out. It is expected to be followed later on by real-world examples that are sufficiently similar to the imaginary situation thus described.

(h) However, (4) is also based on two kinds of intuition: semantic and pragmatic intuition. As regards the former, the thought experiment would not work if the readers' and the authors' *semantic intuition* with respect to the meaning of the English words *argue* and *war* were not made use of. Similarly, the notion of *discourse form,* which is crucial in the inferences (8) and (9), has not been explicated and its interpretation is based on the reader's semantic intuition as well. What is more, the p-context of the thought experiment includes a series of metaphorical expressions which have been mentioned in Lakoff and Johnson (1980: 4) and which motivate the thought experiment. As regards *pragmatic intuition*, Meibauer (2012: 772) points out in general that in reflecting on an imaginary situation, one compares it to a set of contexts that one remembers and that have some properties common with the imaginary situation. Indeed, this is precisely the case with (4)(a), which would not work without this kind of pragmatic intuition. Via these two kinds of intuition, the *second function* of the thought experiment is to *test* Lakoff and Johnson's main hypothesis (3) and thus to contribute to the rejection of the rival claim (2) in Thomason's (1991: 253) sense. As quoted in Subsection 1.3, according Thomason introspective thought experiments serve to test hypotheses about language structure.

(i) Thus, against the background of (g) and (h) above, the thought experiment in (4) is *both* a stage setting *and* an introspective thought experiment. Given Thomason's (1991) rigid distinction between these two kinds of thought experiments, this finding is anything but trivial.

(j) In addition, *the third function* of (4) is to serve as *evidence* for Lakoff and Johnson's main claim (3) and against its rival (2). However, Thomason uses the notion of 'evidence' in a pre-explicative way. In contrast, the p-model defines three versions of evidence (Kertész and Rákosi 2012: 178-185). A datum is weak evidence for a hypothesis if it assigns a positive plausibility value to it. Relative evidence assigns a higher plausibility value to a hypothesis than to its rivals. Strong evidence makes exclusively a given hypothesis plausible and does not support its rivals at all. Obviously, the thought experiment in (4) is *relative evidence* for (3) and against (2).

The next step of our line of reasoning will be another case study, namely, one that is expected to reveal the relationship between the thought experiment in (4) and real experiments. We expect that revealing this relationship will also contribute to the solution of our problem (P).

5. Case study 2: The argumentation structure of real experiments in conceptual metaphor research

5.1. Preliminary remarks

Real experiments as they are carried out in the laboratory are not directly accessible to the scientific community outside the laboratory. It is the *experimental report* that is publicly accessible. Reports on real experiments should also be reconstructed as processes of plausible argumentation (Kertész and Rákosi 2012, 2014, Rákosi 2012). In this way, the relationship between thought experiments and real experiments can be examined within the framework of the p-model. We choose Gibbs et al. (2004) as the subject matter of the case study that will illustrate how the plausible argumentation process initiated by the thought experiment in (4) is

related to real experiments. In the light of our framework, one of the tasks of the real experiments reported on by these authors was to re-evaluate retrospectively one of the consequences of Lakoff and Johnson's thought experiment in (4).

For lack of space, we will dispense with the reconstruction of Gibbs et al.'s (2004) whole argumentation. The plausible inferences into which it can be decomposed are analogous to the mechanism we have exemplified in Section 3, and therefore not all details deserve to be analyzed here. We will focus on those of its aspects that are instructive with respect to our main problem (P).

Furthermore, the reconstruction of the argumentation structure of Gibbs et al.'s experimental report does not follow the linear order in which particular pieces of information have been discussed. The reason is that there is no one-to-one correspondence between the text and the argumentation structure of the experimental report. For example, (16) below is mentioned in the last section of the paper, although it belongs to the starting p-context motivating the experiments that are expected to increase the plausibility of (14).

5.2.   The starting p-context

Lakoff and Johnson's main hypothesis (3), which had been obtained by the thought experiment in (4), gave rise to a series of further tenets. The starting p-context of the experimental report in Section 4 of Gibbs et al. (2004) includes – partly explicitly, partly as latent background assumptions – among others the tenets inferred in the course of the plausible argumentation process initiated by the thought experiment in (4).

As is well known, one of the tenets inferred from (3) in the literature on conceptual metaphors is the following (Gibbs et al. 2004: 1192):

(14)   "Many source domains of conceptual metaphors reflect significant patterns of bodily experience."

In the literature, a complex process of plausible argumentation leads from (3) to (14). This complicated chain of inferences can be subsumed under the pattern of a plausible *modus ponens*: $0<|$If (3), then $(14)|_{SCMR}<1$; $0<|(3)_{S(4)}<1$; therefore, $0<|(14)|_I<1$. (*SCMR* stands for conceptual metaphor research and *S(4)* for the thought experiment in (4).)

Nevertheless, for Gibbs et al., the initial p-context of their research is p-problem*atic* in the sense of (vii) and (viii) in Section 2, because it is informationally underdetermined. Obviously, the authors are not satisfied with (14) as discussed by Lakoff and Johnson and others. The plausibility value of (3) – from which (14) was inferred – is evaluated as considerably lower than it was as the result of (12)(e), i.e. it *decreases* in the starting p-context of the experimental report. The reason is that the authors consider thought experiments to be an *unreliable source* that they contrast with the reliability of real experiments.

So, in order to decide whether (14) is plausible or not, they introduce the conceptual metaphor DESIRE IS HUNGER into the starting p-context and put forward the following hypothesis:

(15)   The more prominent parts of the experimentees' hunger experiences are invariantly mapped onto their different concepts for desire.

If (15) turned out to be plausible, then it would support (14) in the course of the plausible argumentation process through a series of intermediate argumentation steps that can be

subsumed under the following pattern: $0<|$If (15), then $(14)|_G<1$; $0 < |(15)|_{??}<1$; therefore, $0<|(14)|_I <1$. Here G stands for Gibbs et al.'s argumentation as the source of the statement at issue. However, at this point there is no source that could support (15), and this is indicated by the question marks.

Should the authors be able to assign a plausibility value to (15) by finding sources that they assume to be more reliable than Lakoff and Johnson's thought experiments, then, with respect to their main hypothesis in (3), the structure of the inference under which the subsequent chain of further plausible inferences can be subsumed, would be that of *reduction*: $0 < |$If (3), then $(14)|_{SCMR} <1$; $0 < |(14)|_{SCMR} <1$; therefore, $0 < |(3)|_I <1$.[19] Thus, since (14) supports (3) and (14) is supported by (15), the source of which is missing, Gibbs et al.'s argumentation process is expected to find out whether it is possible to identify reliable sources that assign a plausibility value to (15).

The methodological principles in (16)(b) and (c) belonging to the starting p-context constitute such a reliable source:

(16)　(a)　„Does metaphor reflect transcendent metaphysical concerns or is it grounded in embodied experience?
　　　(b)　Our main argument is that *philosophical speculation is not enough* to answer this question.
　　　(c)　*Empirical research is needed* to establish connections between embodiment and metaphor in thought and language." (Gibbs et al. 2004: 1207; emphasis added)

(16) gives rise to the plausible inference in (17):

(17)　Premises:
　　　(a)　[$0<|$If the source of the plausibility of a hypothesis is not empirical research, then it is less plausible than if its source were empirical research.$|_{S(16)(b),(c)} <1$]
　　　(b)　[$|$The source of the plausibility of (14) is not empirical research.$|_{S(16)(b),(c)}<1$]
　　　Conclusion:
　　　(c)　$0<|$The hypothesis in (14) is less plausible than if its source were empirical research.$|_{I(17)}<1$

Another statement the plausibility of which is provided by the same direct source is not explicitly mentioned but is included in the starting p-context as a latent methodological background assumption:

(18)　Empirical research makes use of real experiments.

Accordingly:

(19)　Premises:
　　　(a)　[$0<|$Empirical research makes use of real experiments.$|_{S(16)(b),(c)}<1$]
　　　(b)　$0<|$Empirical research is needed to establish connections between embodiment and metaphor in thought and language.$|_{S(16)(b),(c)}<1$]

---

[19] The pattern of reduction is: $0<|$If A, then $B|_S<1$; $0<|B|_S<1$; therefore, $0<|A|_I<1$. 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).

Conclusion:

(c) $0<|$Real experiments are needed to establish connections between embodiment and metaphor in thought and language.$|_{I(19)}<1$

5.3. *The elaboration of the p-context*

In the first subcycle of the elaboration of the p-context, the authors first examine systematic language patterns expressing desire both in American English and in Brazilian Portuguese. A list of examples including, among others, the following expressions was introduced into the p-context (Gibbs et al. 2004: 1199-1200):

(20) (a) American English:
He hungers for recognition. – He thirsts for recognition.
He hungers for adventure. – He thirsts for adventure.
(b) Brazilian Portuguese:
Tenho fome de riqueza. – Tenho sede de riquezas.
('I hunger for wealth.' – 'I thirst for wealth.')
Tenho fome de saber. – Tenho sede de saber.
('I hunger for knowledge.' – 'I thirst for knowledge.')

The list of examples was randomly collected and their interpretation rests solely on the *linguistic intuition* of the authors. Thus, the following plausible inference can be reconstructed. Thereby, in the subscripts *SI* refers to linguistic intuition as the direct source of the plausibility of the statements at issue.

(21) Premises:
(a) $0<|$If the items in (20)(a) and (b) are expressions of English and Brazilian Portuguese, respectively, then "experiences of hunger appear to structure significant aspects of the speakers' understandings of desire".$|_{SI}<1$ (Gibbs et al. 2004: 1200).
(b) $0<|$The items in (20)(a) and (b) are expressions of English and Brazilian Portuguese, respectively.$|_{SI}<1$
Conclusion:
(c) $0<|$Experiences of hunger appear to structure significant aspects of the speakers' understandings of desire.$|_{I(21)}<1$

The second subcycle was based on the introduction of *a new source*, namely, Experiment 1 in which the experimentees' embodied experiences of hunger apart from their understanding of hunger in talk of desire were investigated. The experimentees (American and Brazilian students) got a random list of symptoms that were assumed to result from somebody's being hungry. The symptoms were divided into three categories: local symptoms, general symptoms and behavior symptoms. Each of these categories was further subdivided into closely related, possibly related and not related symptoms. The experimentees' task was to determine on a 7-point scale if they had experienced a particular effect on the list when feeling hungry. The result was similar in the case of the English speaking and the Portuguese speaking experimentees, because they gave similar ratings to the different items (see Gibbs et al. 2004: 1202-1203):

(22)   Strong effects of hunger on the human body:
   (a)   Local symptoms: the stomach grumbles, thought of food makes one's mouth water, one has a stomachache, one has a headache. English: 5.39. Portuguese: 5.78
   (b)   General symptoms: feel discomfort, becomes weak, become dizzy, gets annoyed, have an appetite. English: 6.03. Portuguese: 6.12.
   (c)   Behavior symptoms: person feels out of balance, person becomes emotionally fragile, person becomes very anxious. English: 4.74. Portuguese: 5.09.

The results were also similar with respect to those items that were not related to the experimentees' hunger experiences:

(23)   Weak effects of hunger on the human body:
   (a)   Local symptoms: the knees swell, the feet hurt, the hands itch, the fingers snap. English: 2.02. Portuguese: 1.88.
   (b)   General symptoms: wants to run, doesn't wish to see anyone, becomes talkative, gets a fever. English: 3.27. Portuguese: 2.84.
   (c)   Behavior symptoms: the person behaves normally, the person can work well. English: 2.63. Portuguese: 2.26. (Gibbs et al. 2004: 1203)

The plausible inference built on these findings is this:

(24)   Premises
   (a) $[0<|$If (22) and (23), then there are "significant regularities in people's embodied experiences of hunger".$|_{SE1}<1]$ (Gibbs et al. 2004: 1204)
   (b) $0<|$(22) and (23)$|_{SE1}<1$
   Conclusion:
   (c) $0<|$There are significant regularities in people's embodied experiences of hunger.$|_{I(24)}<1$

In the authors' view, (24)(c) provided "a rough characterization of folk understandings of embodied experiences for hunger" (Gibbs et al 2004: 1204).

In (24)(a) and (b) the subscript *SE1* refers to Experiment 1 as the direct source in which the plausibility of the statement is rooted. Our analysis via the p-model suggests that, as a result of this subcycle, (24)(c) *retrospectively re-evaluates* (21)(c) by providing a new source that increases its plausibility value. At the same time, it adds new information to the p-context that might be used in the next stage of the argumentation process. In addition, the plausible argumentation process is *prismatic* as well, because (21)(c) is re-evaluated from a new perspective, namely, that of Experiment 1, as a new direct source.

Then, in the third subcycle, the p-context was further extended by *another direct source*: Experiment 2. Its aim was to find out if folk understanding about hunger as characterized in (24)(c) is correlated with how people understand different experiences of desire. The experiment consisted of two sets of questions to be answered by the experimentees (who were not identical with those of the first experiment). With the help of the first set of questions, the experimentees were asked how they felt when they experienced love, lust and a series of further desires such as desire for fame, money etc. Half of the questions focused on body experiences that during the first experiment were rated as being of high relevance to one's experiences of hunger. The other half made use of body experiences that were rated as being of low relevance to hunger experiences. For example:

(25)   "How do you imagine that somebody who is deeply in love feels?
        the person becomes dizzy
        the person becomes weak
        the person gets annoyed
        the person becomes talkative" (Gibbs et al 2004: 1205)

After reading the question at issue, participants had to rate the relevance of different bodily experiences on a 7-point scale.

       With the help of the second set of questions within the second experiment, experimentees were asked to rate on a 7-point scale whether the given statement was *acceptable* in their respective language i.e. English or Brazilian Portuguese. Half of the items were chosen from those that were highly rated in the first experiment, and the other half was compiled from weakly rated items. The questions focused on the same three types of desire as in Experiment 1. The findings were as follows (also summarized in Table 1 of the experimental report):

(26)   "The findings for both the Body and Linguistic questions are generally consistent across English and Portuguese for the three types of symptoms for the three types of desire (love, lust, other). Each difference between the strong and weak items for each type of desire was statistically significant, with the exception of love and other desire for English speakers which were only marginally different." (Gibbs et al. 2004: 1206)

In the view of the authors, these data "demonstrate how knowing something about people's embodied experiences of hunger allows us to predict which aspects of desire will, and will not, be thought of and talked about in terms of our complex embodied understandings of hunger" (ibid.). Accordingly, the structure of the plausible inference resting on Experiment 2 is as follows:

(27)   Premises
        (a) $[0<|$If (26), then the more prominent parts of the experimentees' hunger experiences are invariantly mapped onto their different concepts for desire.$|_{SE2}<1]$
        (b) $0<|(26)|_{SE2}<1$
        Conclusion:
        (c) $0<|$The more prominent parts of the experimentees' hunger experiences are invariantly mapped onto their different concepts for desire. $|_{I(27)}<1$

In the subscripts *SE2* stands for Experiment 2 as the direct source of the plausibility of the statements. (27)(c) *retrospectively and prismatically re-evaluates* (24)(c) and (21)(c).

5.4.   The final p-context

(21)(c), (24)(c) and (27)(c) yield the hypothesis (15). This means that now the latter can be assigned a higher plausibility value than in the starting p-context. The reason is that its plausibility was provided by argumentation cycles making use of *four direct sources*: methodological principles, introspection and two experiments.

       However, it is a solution, but no resolution of the p-problem raised in the initial p-context – that is, whether (14) is plausible or not. Namely, it is one of the premises of further

plausible inferences constituting a new subcycle of the plausible argumentation process. Consider:

(28)   (a)   "If hunger and desire are highly correlated, and if people metaphorically make sense of their desires partly in terms of hunger,

      (b)   then these more prominent parts of their hunger experiences should be invariantly mapped onto their different concepts for desire." (Gibbs et al 2004: 1200)

The structure of the inference inherent in this quotation is *reduction*:

(29)   Premises:

      (a) $0<|$If hunger and desire are highly correlated, and if people metaphorically make sense of their desires partly in terms of hunger, then these more prominent parts of their hunger experiences should be invariantly mapped onto their different concepts for desire. $|_{SCMR}<1$

      (b) $0<|$These more prominent parts of their hunger experiences are invariantly mapped onto their different concepts for desire.$|_{I(27)}<1$

   Conclusion:

      (c) $0<|$Hunger and desire are highly correlated, and people metaphorically make sense of their desires partly in terms of hunger.$_{I(29)}<1$

The plausibility of (29)(b) rests on an indirect source, namely, (27). However, (27)(c) has been inferred, among others, from (27)(b) whose plausibility is rooted in a direct source, namely the second real experiment. The direct source of the plausibility of (29)(a) is the state of the art of conceptual metaphor research (*CMR*) as initiated by (4). Then, the next inference is reconstructed in (30):

(30)   Premises:

      (a) $0<|$If hunger and desire are highly correlated, people metaphorically make sense of their desires partly in terms of hunger, and hunger is a source domain of a conceptual metaphor that reflects significant patterns of bodily experience, then many source domains of conceptual metaphors reflect significant patterns of bodily experience. $|_{SCMR, I(29)}<1$

      (b) $0<|$Hunger and desire are highly correlated, and people metaphorically make sense of their desires partly in terms of hunger.$|_{I(29)}<1$

      (c) $0<|$Hunger is a source domain of a conceptual metaphor that reflects significant patterns of bodily experience.$|_{SEI}<1$

   Conclusion:

      (d) $0<|$Many source domains of conceptual metaphors reflect significant patterns of bodily experience.$|_{I(30)}<1$

Here the premise (30)(b) is identical to the conclusion of (29), i.e. (29)(c), and its plausibility is rooted in an indirect source, i.e. the inference (29). In turn, (29)(c) has been obtained, among other things, from the premise (29)(b). Therefore, the conclusion (30)(d) has been obtained indirectly with the help of (29)(b). However, (29)(b) is identical to (15), and (30)(d) is identical to (14). Thus, (14) has been supported indirectly by (15) during the plausible argumentation process. In Lakoff and Johnson's theory, (14) had not been based on real experiments, whereas in Gibbs et al. (2004), the plausibility value of (15) was obtained by an argumentation process that among others made use of real experiments as sources of the

plausibility of statements. Consequently, the feedback mechanism increased the plausibility of an originally non-experimentally based plausible statement, i.e. (14), on the basis of real experiments carried out in a later phase of the argumentation process.

Then, the final step follows. Its structure is *reduction* as already mentioned in Subsection 5.2:

(31)   Premises:
        (a) $0 <$|If our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature, then many source domains of conceptual metaphors reflect significant patterns of bodily experience.$|_{SCMR} < 1$
        (b) $0 <$| Many source domains of conceptual metaphors reflect significant patterns of bodily experience.$|_{I(30)} < 1$
      Conclusion
        (c) $0 <$|Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature.$|_{I(31)} < 1$

Obviously, (31)(b) corresponds to (15) i.e. the main hypothesis of Gibbs at al. (2004). In (31), its plausibility rests on an indirect source, namely, (30). However, via (24), (27) and (29), it was inferred as the result of a chain of inferences including premises whose plausibility was based on Experiments 1 and 2 as direct sources. In this way, the fact that in the final p-context of the real experimental report the plausibility value of (15) has increased relative to that in the starting p-context, contributes to the retrospective re-evaluation of (3) (which is equivalent to (31)(c)) and was originally obtained by the thought experiment in (4)) – thus increasing the latter's plausibility.

Now, we must not forget that, as we have seen in Section 3, (12)(e), and through this, (3) – that is, Lakoff and Johnson's main hypothesis – were obtained as the result of a plausible argumentation process that constituted a thought experimental report. Therefore, (29), (30) and (31) work as prisms through which the outcome of Lakoff and Johnson's thought experiment is re-evaluated retrospectively. It is important to emphasize that this process, as a result of which the finding of Gibbs et al.'s real experiments feeds back to the final p-context of Lakoff and Johnson's thought experiment and increases its plausibility, is not circular.[20] Rather, it is *a clear case of the cyclic, retrospective and prismatic re-evaluation of information* in the sense of the p-model.

6.   *Summary of the findings of case study 2*

Within the overall process of plausible argumentation, the relationship between the thought experiment and the real experiments as reported on has the following features:

(a) The case study has shown that real experimental reports are plausible argumentation processes. The *starting p-context* includes the results of the argumentation process for which the stage was set by the thought experiment reported on in (4) (see also point (g) in Section 4). However, the result of the thought experiment was integrated into the starting p-context of the real experimental report in a modified form, because its plausibility value *decreased*. Consequently, there is a *first kind* of indirect *continuity* between the thought experiments and the real experiments, but this continuity is *not linear*; rather, it is cyclic, prismatic and retrospective.

---

[20] For the distinction between circular and cyclic argumentation in conceptual metaphor research, see Kertész and Rákosi (2009).

(b) The *result* of Gibbs et al.'s real experiments retrospectively re-evaluated the result of Lakoff and Johnson's thought experiment again. This time they *increased* the latter's plausibility as the outcome of the argumentation process based on the real experiments as direct sources of some of the premises on which particular plausible inferences constituting this process were built. Accordingly, there is *a second kind* of cyclic, prismatic and retrospective continuity between the thought experiment and the real experiments *in the opposite direction* as well.

(c) The literature on thought experiments in the natural sciences highlights the fact that they anticipate real experiments: in order to design a real experiment, first it has to be conducted in thought (for detailed discussion see Buzzoni 2008; see also Kertész and Rákosi 2014 with respect to experimental pragmatics). This applies to the real experiments we discussed, too, and corresponds to a *third kind* of continuity between thought experiments and real experiments in conceptual metaphor research. In Thomason's (1991) terminology, this means that stage setting thought experiments anticipate real experiments.

(d) The experiments had introspective components. First, as already emphasized, in order to anticipate the experiments, Gibbs et al. enumerate expressions such as in (20) which were simply gathered on the basis of the authors' native speaker knowledge of American English and Brazilian Portuguese – that is, by introspection, and it was assumed that there is a conceptual metaphor DESIRE IS HUNGER underlying them. Second, when in the first experiment the experimentees had to rate the items according to the scale mentioned with respect to the question whether they had experienced the effects listed when feeling hungry, the only source of their answers was introspection. Accordingly, both the experimenters and the experimentees carried out thought experiments based on introspection. Finally, in the second experiment „the set of questions focused on *participants' intuitions* about the acceptability of different ways of linguistically expressing desire" (Gibbs et al 2004: 1205; emphasis added). Consequently, if – as Thomason (1991) claims – introspection is a kind of thought experiment, then the real experiments *include* components that correspond to thought experiments. However, without these components the real experiments would not have worked. Therefore, in this sense, real experiments and thought experiments are closely *interwoven*.

7.   *On (P)*

7.1.   Summary

For the sake of clarity, let us overview the steps of our reasoning before we proceed to the solution of (P):

Step 0: Raising the problem (Section 1).

Step 1: Introduction of the research framework (Section 2): The p-model has been introduced as the research framework the application of which should lead to the solution of (P).

Step 2: The application of the research framework to the analysis of a well known thought experiment in conceptual metaphor research (Section 3). The p-model as our research framework has been applied to the analysis of the thought experiment in (4).

Step 3: Drawing conclusions from the application of the research framework to the thought experiment in (4) (Section 4). The application of the p-model to the thought experiment at issue has revealed some of its *structural* and *functional* properties. These properties present themselves as *conclusions* yielded by the analysis as carried out in Section 3.

Step 4: The application of the research framework to the analysis of real experiments in conceptual metaphor research (Section 5). Now, presupposing what the application of the framework has already revealed about the structure and the function of thought experiments in conceptual metaphor research (as exemplified by the thought experiment in (4)), the p-model has been applied to the analysis of real experiments.

Step 5: Drawing the conclusions from the second case study (Section 6). The second case study has shown, among others, that the real experiments *cannot work* without the thought experiment and their properties are affected by the structural and functional properties of the thought experiment. Conversely, the acceptability of the findings that the thought experiment yields, crucially depends on the findings of the real experiments etc.

Step 6: Drawing the conclusions from the previous findings. The solution to (P) should be obtained from the findings of the above line of reasoning.

We have already carried out steps 0-5. Accordingly, the next subsection has to be devoted to step 6.

7.2.   The solution to (P)

The conclusions we have drawn from the two case studies lead to the solution to (P) in the following way:

(a) In (13), we maintained that Lakoff and Johnson's thought experiment included an analogical inference that connected the imaginary situation described in (4)(a) with real communicative situations one has already experienced. Accordingly, (13) and point (f) in Section 4 suggest that the conclusion of such an analogical inference contains *empirical information about some properties of particular kinds of personal experiences with real situations*. This information has already been included in the latent background assumptions. Nevertheless, it is important to realize that the sources that support such latent background assumptions – i.e. personal experiences of the reader – are part of the p-context but are not part of the thought experiment. Therefore, in this case, the thought experiment contributes indirectly to adding this new empirical information to the plausible argumentation process by means of such an analogical inference.

(b) According to point (b) in Section 6, the new empirical information obtained as the result of the real experiment may retrospectively re-evaluate the result of the thought experiment already included in the starting p-context of the real experiment. Therefore, the real experiment is the *empirical source* of the plausibility of the re-evaluated hypothesis that the *thought experiment* yields. The outcome of this retrospective re-evaluation of the hypothesis originally obtained from the thought experiment may be empirical, but merely in an indirect way, *through the prism of the real experiment*.

(c) In Section 4 point (c) maintains that among the plausible inferences into which a thought experiment in conceptual metaphor research can be decomposed, there are enthymematic ones. As we know, in such inferences – by means of the latent background assumptions – the conclusion may contain more information than the premises. This new information *may, but need not,* be empirical. If it is empirical, then the conclusion, and via this, the result of the thought experiment, *may contain new empirical information* that has not been put into the explicit premises. In this case, the source which supports the plausibility of a given latent background assumption is included in the p-context, but is external to the thought experiment. So, the thought experiment can transmit empirical information to the argumentation process indirectly, via the latent background assumption that is contained in the p-context and is used in an enthymematic plausible inference (see also Kertész and Kiefer 2013 on this).

We have not disputed that thought experiments in conceptual metaphor research might contribute to new empirical insights. Our main problem (P) has been *how* a thought experiment in conceptual metaphor research leads to novel empirical information. At this point of our line of reasoning, we can infer the solution of (P) that we have already indicated in Section 1 and that we repeat here (in a slightly more technical manner) for convenience:

(SP) (a)   Thought experiments in conceptual metaphor research are to be conceived of as complex phases in the dynamic process of plausible argumentation.

        (b)   In accordance with this, new empirical information is obtained from sources which are outside the thought experiment, but which have been put into the p-context by the continuous, cyclic, prismatic and retrospective process of plausible argumentation: for example, personal experiences with real world situations (see (a) above) or real experiments (see (b) above).

        (c)   Thought experiments may transmit empirical information from the p-context to the particular cycles of the process of plausible argumentation (see (c) above).

        (d)   Consequently, a thought experiment cannot provide new empirical information itself, but the novel insights it yields consist in its contribution to the cyclic, prismatic and retrospective re-evaluation of empirical information.

        (e)   Thereby, the already known knowledge put into this process is not identical with its output, because in the course of the cyclic, prismatic and retrospective re-evaluation, the plausibility value of the statements at issue *changes dynamically.*

As already mentioned in Section 1, the most important novelty of this solution to (P) is due to the fact that it accounts for the nature of thought experiments in conceptual metaphor research by characterizing them as constitutive components of the dynamic process of plausible argumentation.

7.3.   *The utility of the solution to (P)*

In conceptual metaphor research, there are basically two antagonistic views with respect to the use of thought experiments and real experiments. The first says that thought experiments are not empirical, thus unsatisfactory and should therefore be replaced by real experiments (see e.g. the quotation in (16) and the latent background assumption in (18)). The second maintains the opposite: namely, that thought experiments still work well and should be retained. (For contrasting these two views see e.g. Gibbs 2007, Itkonen 2003). Consequently, *the utility* of the solution to (P) is that it legitimizes a much more sophisticated stance: their integration. It suggests that in conceptual metaphor research, both of them should be applied in accordance with the mechanism of the plausible argumentation process as characterized by the p-model. This cyclic, retrospective and prismatic integration of thought experiments and real experiments may increase the effectiveness of conceptual metaphor research. This is an insight with immediate consequences for the implementation of the everyday practice of inquiry in this field.

The present paper has witnessed that the p-model may provide an original solution to one of the problems that has been in the focus of current discussions. However, it will be the task of further case studies to clarify whether, and if so, in what way and to what extent, our solution to (P) can be refined in order to capture the wide range of thought experiments used in conceptual metaphor research, the varieties of their interaction with real experiments, and

what is more, to examine whether it can be extended to thought experiments in other fields of research.[21]

Acknowledgement

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. 2011. Thought experiments. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/thought-experiment/. Accessed 12th July 2013.

Buzzoni, Marco. 2008. *Thought Experiment in the Natural Sciences: An Operational and Reflexive-Transcendental Conception*. Würzburg: Königshausen and Neumann.

Cooper, Rachel. 2005. "Thought experiments." *Metaphilosophy* 36: 328-347.

Cohnitz, Daniel. 2006. *Gedankenexperimente in der Philosophie*. Paderborn: Mentis.

Galileo. 1974. [1638]. *Two New Sciences*. (translated from the *Discoursi* by S. Drake). Madison: University of Wisconsin Press.

Gendler, Tamar Szabó. 2000. *Thought Experiments: On the Power and Limits of Imaginary Cases*. New York: Garland Publishing.

Gibbs, Raymond W. Jr. 2007. Why cognitive linguists should care more about empirical methods. In *Methods in Cognitive Linguistics,* ed. by Mónica Gonzáles-Márquez, Irene Mittelberg, Seana Coulson, and Michael J. Spivey, 2-18. Amsterdam and Philadelphia: Benjamins.

Gibbs, Raymond W. Jr., Paul Lenz Costa Lima, and Francoso, Edson. 2004. Metaphor is grounded in embodied experience. *Journal of Pragmatics* 36: 1189-1210.

Gooding, David. 1990. *Experiment and the Making of Meaning*. Dordrecht: Kluwer.

Häggquist, Sören. 1996. *Thought Experiments in Philosophy*. Stockholm: Almquist and Wiksell International.

Horovitz, Tamara and Gerald J. Massey. 1991. Introduction. In Horovitz and Massey (eds), pp. 1-28.

Horovitz, Tamara and Gerald J. Massey (eds) 1991. *Thought Experiments in Science and Philosophy*. Savage MD: Rowman and Littlefield.

Itkonen, Esa. 2003. *What is Language? A Study in the Philosophy of Linguistics*. Turku: University of Turku.

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. Berlin, Heidelberg and New York: Springer, 53-86.

---

[21] Thereby, due to the great variety of thought experiments, one must be very cautious in order to avoid hasty generalizations. For example, with respect to thought experiments in pragmatics, our solution to (P) is analogous to that in Kertész and Kiefer (2013) and applies to certain thought experiments in pragmatics as well. However, the analysis of another pragmatic thought experiment and further real experiments in Kertész and Rákosi (2014) is different to some extent.

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. 2014. Thought experiments and real experiments as converging data sources in pragmatics. In András Kertész and Csilla Rákosi (eds), 221-269 (in press).

Kertész, András, Monika Schwarz-Friesel and Manfred Consten (eds). 2012. *Converging Data Sources in Cognitive Linguistics.* Special Issue of *Language Sciences* 34/6, 651-778.

Kertész, András and Csilla Rákosi (eds). 2014. *The Evidential Basis of Linguistic Argumentation*. Amsterdam and Philadelphia: Benjamins.

Kuhn, Thomas S. 1977. A function for thought experiments. In *The Essential Tension*: *Selected Studies in Scientific Tradition and Change*, by Thomas S. Kuhn, 240-265. Chicago: The University of Chicago Press.

Kühne, Ulrich. 2005. *Die Methode des Gedankenexperiments*. Frankfurt am Main: Suhrkamp.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, George and Mark Johnson. 1999. Philosophy in the flesh. The embodied mind and and its challenge to Western thought. New York: Basic Books.

Łukasiewicz, Jan. 1970 [1912]. Creative elements in science. In *Selected Works,* by Jan Łukasiewicz, 12-44. Amsterdam: North Holland.

McAllister, James. 1996. "The evidential significance of thought experiment in science." *Studies in History and Philosophy of Science* 27: 233-250.

Meibauer, Jörg. 2012. Pragmatic evidence, context, and story design. An essay on recent developments in experimental pragmatics. In Kertész et al. (eds), 768-776.

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.

Miščević, Nenad. 1992. "Mental models and thought experiments." *International Studies in the Philosophy of Science* 6: 215-226.

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.

Peijnenburg, Jeanne and David Atkinson. 2003. When are thought experiments poor ones? *Journal for General Philosophy of Science* 34: 305-322.

Polya, George. 1948. *How to Solve It*. Princeton: Princeton University Press.

Polya, George. 1954. *Patterns of Plausible Inference*. London: Oxford University Press.

Rákosi, Csilla. 2012. The fabulous engine: strengths and flaws of psycholinguistic experiments. In Kertész et al. (eds), 682-701.

Rapp, Christof. 2010. Aristotle's rhetoric. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/aristotle-rhetoric/. Accessed 12th July 2013.

Rescher, Nicholas. 1976. *Plausible Reasoning*. Assen and Amsterdam: Van Gorcum.

Rescher, Nicholas. 1987. How serious a fallacy is inconsistency? *Argumentation* 1: 303-316.

Sorensen, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.

Thomason, Sarah G. 1991. Thought experiments in linguistics. In Horovitz and Massey (eds), 247-257.

Walton, Douglas N. 1992. *Plausible Argument in Everyday Conversation*. Albany: State University of New York Press.

Walton, Douglas N., Chris Reeds and Fabrizio Macagno. 2008. *Argumentation Schemes*. New York: Cambridge University Press.

Author Biography

András Kertész is Professor of Linguistics at the University of Debrecen and Head of the Research Group for Theoretical Linguistics of the Hungarian Academy of Sciences. In 2001 he was elected as a member of the Hungarian Academy of Sciences. His main fields of research are the philosophy of linguistics and theoretical linguistics. He published 10 monographs, 9 collections of papers and about 130 articles. His books include *Cognitive Semantics and Scientific Knowledge* (Amsterdam: Benjamins, 2004), *Philosophie der Linguistik* (Tübingen: Narr, 2004), *Data and Evidence in Linguistics: A Plausible Argumentation Model* (Cambridge: Cambridge University Press, 2012; co-author: Csilla Rákosi) and *The Evidential Basis of Linguistic Argumentation* (Amsterdam: Benjamins, 2014, co-editor: Csilla Rákosi).

http://web.t-online.hu/andraskertesz/angol.htm