András Kertész: The armchair in the laboratory
Argumentum 10 (2014), 359-371
Debreceni Egyetemi Kiadó

András Kertész

The armchair in the laboratory

A note on the relationship between introspective thought experiments and real experiments in linguistics

Abstract

The present note is meant to be a modest contribution to the current discussion on the nature of linguistic data. It focuses on the question of what kind of relationship there is between introspective thought experiments and real experiments in linguistics. In order to answer this question, Kertész and Rákosi's (2012) p-model is introduced as a metatheoretical framework and applied to two examples: to Jackendoff's (1994) treatment of grammaticality judgments and Gibbs et al.'s (2004) real experiments on conceptual metaphor. These two examples illustrate the main finding that says that there is a multifaceted interaction between introspective thought experiments and real experiments in linguistics: thought experiments may be components of real experiments, real experiments are motivated by introspective thought experiments, and the results of real experiments retrospectively re-evaluate those of introspective thought experiments.
Keywords: thought experiment, real experiment, conceptual metaphor research, grammaticality judgment, plausible argumentation

1 Introduction

In the course of the controversy on the nature of linguistic data,[1] introspective data have been evaluated in three different ways. The first view, which has primarily been advocated by generative linguists from the late fifties on, says that grammaticality judgments gained by introspection are (a) fully legitimate, (b) empirical, and (c) maximally reliable.[2] They have often been characterized in analogy to experiments used in the natural sciences. As an example, let us take the following passage from a textbook designed to convey generally accepted knowledge to the public:[3]

(1) "What is a linguistic experiment? As in other sciences, the strategy is to study unobservable phenomena by relating them to things that *are* observable. […] **The only**

[1] For a detailed overview of the data/evidence problem in linguistics, see Kertész & Rákosi (2012).

[2] For the criticism of the notions of instrospection and intuition see Schütze (1996), Itkonen (2003) and Schütze & Sprouse (2013).

[3] There are a great number of similar formulations in Chomsky's work as well. Here is an example:
"In actual practice, linguistics as a discipline is characterized by attention to certain kinds of *evidence* that are, for the moment, readily accessible and informative: largely, *the judgements of native speakers*. Each such judgement is, in fact, *the result of an experiment*, one that is poorly designed but rich in the *evidence* it provides" (Chomsky 1986: 36).

> **difference is that linguistic experiments have to do with the inside of our heads** instead of external objects. […]. The idea is that although we can't observe the mental grammar of English itself, **we *can* observe the judgments of grammaticality and meaning** that are produced by using it […]. (Jackendoff 1994: 46; bold emphasis added, italics as in the original)

In contrast, exactly what the above quotation considers to be *conformity* to the experimental method has been interpreted by critics as the *violation* of the methodological norms of experimentation:

(2) "[…] there are important shortcomings that arise because linguistic elicitation does *not* follow the procedures of psychological experimentation. Unlike natural scientists, **linguists are not trained in methods for getting reliable data and determining which of two conflicting data reports is more reliable**. In the vast majority of cases in linguistics, there is **not the slightest attempt to impose any of the standard experiential control techniques**, such as random sampling of subjects and stimulus materials or counterbalancing for order effects. […] Perhaps worst of all, often the only subject of these **pseudoexperiments** is none other than the theorist himself or herself […]. (Schütze 1996: 4; bold emphasis added, italics as in the original)

Some generative linguists seem to accept this criticism and go even so far as to require the banishment of introspective data from theoretical linguistics (Haider 2009).

The third position says that introspective data alone are insufficient, but, if combined with other data types, they may be useful and even indispensable:

(3) "[…] intuition should not be condemned as an inadequate source of data and pushed aside completely. Even linguists who work experimentally and use empirical corpus data cannot ignore the introspective method. Qualitative analysis is always based on some mental process of construction and formulation of hypotheses relying on intuition. […] *What is needed is a combination of both introspectively gathered data and empirically based evidence*." (Schwarz-Friesel 2012: 660; emphasis added)

The aim of the present note is to argue for the standpoint exemplified in (3) from a specific point of view which, to our knowledge, has not been systematically investigated so far: it concerns the relationship between introspective thought experiments and real experiments in linguistics. Consequently, the problem we will tackle is this:

(P) What kind of relationship is there between introspective thought experiments and real experiments in linguistics?

In order to answer the question raised in (P), we will proceed as follows:

In Section 2, we will relate data based on introspection to the notion of the thought experiment. In order to solve (P), it is indispensable to introduce an appropriate methodological framework. Therefore, Section 3 will be devoted to sketching the basic ideas of such a framework. Then, we will illustrate the workability of this framework by two examples. In Section 4, using Jackendoff's (1994) treatment of grammaticality judgments, we will examine (P) from the point of view of introspective thought experiments. In turn, in

Section 5, we will start from real experiments in order to examine the relationship at issue. Finally, in Section 6, we will infer our solution to (P) from the analysis of the two examples.

This note is meant to be a short and simplified contribution to the present discussion on linguistic data without striving for completion and technical precision. For more comprehensive treatments of related topics, see Kertész & Kiefer (2013), Kertész & Rákosi (2014), Kertész (2014).

2 On thought experiments

Although thought experiments have been frequently used as tools of scientific research and philosophy alike, and have also undoubtedly influenced some of the great scientific achievements of the twentieth century, until recently their nature has been a neglected topic in the philosophy of science. However, over the past two decades they have been the subject of a fierce debate (for short overviews, see Brown & Fehige 2013 and Moue et al. 2006). Although thought experiments are widely used in different fields of linguistics, so far they have not been related to this discussion in the philosophy of science and their properties have not been analysed systematically, either. The only survey of thought experiments in linguistics I know of is Thomason (1991), which discusses two kinds of thought experiments in linguistics: stage setting and introspective thought experiments. By stage setting thought experiments, Thomason (1991) means that the thought experiment functions as a first step in the argumentation in which it clarifies the theoretical issue. Introspective thought experiments, however, have a different function:

(4) "Let us turn now to the other kind of linguistic thought experiment – *the kind that involves introspection*, by the linguist or by an informant (a native speaker of some language the linguist is investigating), about the appropriateness of a particular linguistic form or construction. Thought experiments of this type are *actual tests of hypotheses* about language structure." (Thomason 1991: 252-253; emphasis added)

At this point it is sufficient to notice that, if we accept Thomason's notions, grammaticality judgments are to be conceived of as the results of introspective thought experiments.

Our second argument that supports the claim that grammaticality judgments can be conceived of as the outcomes of introspective thought experiments pertains to the quotation in (1). Jackendoff claims that grammaticality judgments as introspective data are the results of experiments and have evidential significance. Thereby, this evidence is only indirect in that it is located in the informants' *heads*. However, the idea that introspection yielding grammaticality judgments works analogously to real experiments, the only difference being that they are performed in one's "head", is identical to the criterion according to which philosophers of science distinguish real experiments from thought experiments. Consequently, Jackendoff commits a category mistake when he blurs this difference and interprets grammaticality judgments as the results of real experiments claiming that "grammaticality judgments remain the most widely used experimental technique in contemporary linguistics" (Jackendoff 1994: 49-50).

András Kertész: The armchair in the laboratory
Argumentum 10 (2014), 359-371
Debreceni Egyetemi Kiadó

3 The methodological framework: the p-model

In order to obtain a possible solution to (P), we will apply Kertész and Rákosi's (2012) *p*-model of plausible inferences and plausible argumentation to the analysis of introspective thought experiments. The p-model is a general approach to the philosophy of linguistics and is not restricted to thought experiments. Its purpose is to yield a novel solution to the data/evidence problem in theoretical linguistics and to provide tools for the analysis of linguistic theory formation. Thereby, it defines a linguistic theory as a dynamic process of plausible argumentation. Below we merely summarise some of its underlying ideas in a highly simplified and informal way. For details, see Kertész & Rákosi (2012).

The main characteristic of *plausible inferences* is that – in opposition to deductive inferences the conclusions of which are true with certainty, provided that the premises are also true – the conclusions of plausible inferences are merely plausible, i.e. uncertain and fallible, although they have heuristic power. In Kertész & Rákosi (2012) three basic types of plausible inference are considered. Inferences belonging to the first type include at least one premise that is not true but only plausible. Therefore, the conclusion is also merely plausible. In inferences of the second type – called enthymematic – the premises may be true, but there is no logical consequence relation between them and the conclusion. Therefore, *latent background assumptions* work as hidden premises that make the inferences logically valid. Finally, these two types may be combined.

Thus, the conclusion of a plausible inference, and in types 1 and 3 at least one of the premises, is a *plausible statement*. A plausible statement consists of a statement and a plausibility value. Thereby, the plausibility value is provided by the reliability of the *source* it is rooted in. The plausibility value of statement *p* on the basis of the source *S* is such that:

(5) $|p|_S = 1$, 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.

Here plausible statements are set within '|'. Outside of the latter there are the plausibility values assigned to the statement based on a given source.

We assume that thought experiments are heuristic tools of problem solving. A certain phase of theory formation may be *problematic*, in the first case if it is overdetermined by information in the sense that both a certain statement and its negation are present; that is, if it is inconsistent. Second, it may be informationally underdetermined insofar as there are statements that are neither plausible nor implausible. Third, it may be both over- and underdetermined with respect to different statements. We call the tool that serves the elimination of the under- and/or overdetermination – namely, the problem-solving tool –, *plausible argumentation*. It consists of a sequence of plausible inferences and is processual. However, the process of plausible argumentation is not linear. Rather, during the plausible argumentation process one returns to previous phases and *retrospectively re-evaluates* former findings in a *cyclic* (but not circular) way.[4] Another feature of the retrospective re-evaluation of the findings is that it is *prismatic* (in Rescher's 1987 sense) because it is carried out from

[4] For the distinction between cyclic and circular argumentation in cognitive linguistics, see Kertész & Rákosi (2009).

continuously changing perspectives during which new information is considered, or earlier findings are modified, deleted or supplemented by additional assumptions, etc.

In the next section we will apply the p-model in order to reveal certain properties of grammaticality judgments as the results of introspective thought experiments.

4 First example: Grammaticality judgments as results of introspective thought experiments

Up to this point, we have used the notions 'datum', 'introspective datum' and 'introspection' pre-explicatively. Before proceeding, we must now make the following distinctions:

A datum is a plausible statement stemming from a direct source. For example, 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.

Thus, data are, by definition, never certain, but merely plausible. Let us illustrate this general definition of the notion of datum by the specific case of grammaticality judgments:

(6) (a) $0<|$The sentence
It is unclear what shocked whom
is grammatically correct.$|_S<1$
(b) $0<|$The sentence
It is unclear whom what shocked
is ungrammatical.$|_S<1$
(c) *It is unclear what shocked whom.*

The *statements* in (6)(a) and (b) are introspective *data* whose *direct source S* may be the introspection of one or more informants. If one trusts introspection as a source, then (6)(a) and (b) are accorded a high plausibility value. If, in turn, one is a corpus linguist questioning the reliability of introspection, then they will have a low plausibility value. Such statements claim that it is plausible/implausible that the sentence mentioned does or does not have the property of being grammatically correct. Accordingly, the *sentence* itself in (6)(c) does *not* count as a datum.

Since (a) introspection is the direct source of grammaticality judgments, (b) data are plausible statements, and (c) the plausibility of statements depends on the reliability of the source they stem from, the plausibility degree of grammaticality judgments depends on the reliability of introspection. This characterisation of grammaticality judgments is at variance with its classical treatment according to which introspection "is *so reliable* that, for a very good first approximation, linguists tend to trust their own judgments and those of their colleagues" (Jackendoff 1994: 48; emphasis added).

Why do we consider introspection, in opposition to this quotation, as an unreliable source and grammaticality judgments as plausible rather than true statements? The answer can be given easily in the light of the recent literature (see Kertész & Rákosi 2012 and Schütze & Sprouse 2013 for overviews). For example, the p-model interprets the following factors discussed in the latest literature at length as speaking for the unreliability of introspection as a data source and the mere plausibility of grammaticality judgments: the unsystematicity of data

collection; the variation between interspeaker judgments; the graduality of the judgments; the manipulability of judgment data, etc.

We will illustrate the way the p-model treats introspective thought experiments by a very simple example taken from Jackendoff's (1994) classic textbook. The reason why we chose this example is that Jackendoff's textbook typically represents the *almost unrestricted reliance* on introspection in a period in which the present discussion on linguistic data had not yet been raised. Therefore, it will be relatively simple to contrast his view with the very different perspective of the p-model.

Jackendoff raises the initial problem in the following way:

(7) Is there a mental grammar?

Obviously, this phase of theory formation is informationally underdetermined in the sense of the p-model, because neither of the alternative answers to this question can be assigned a plausibility value on the basis of the sources available at this stage of the argumentation process:

(8) (a) $|$There is a mental grammar.$|_S = 0$
(b) $|$There is no mental grammar.$|_S = 0$

First, he has carried out an introspective thought experiment in Thomason's sense. By this introspective thought experiment, he has obtained the following grammaticality judgments whose plausibility stems from his own intuition as a direct source:

(9) (a) $0<|$The sentence *Amy ate two peanuts* is grammatical.$|_{SJ}<1$ (Jackendoff 1994: 11).
(b) $0<|$It is not the case that the sentence Amy two ate peanuts is grammatical.$|SJ<1$ (Jackendoff 1994: 15).

In the subscripts of (9)(a) and (b), *SJ* stands for Jackendoff's intuition as a direct source.

The argumentation Jackendoff (1994: 15-16) carries out starting from the introspective thought experiments in (9)(a) and (b) can be reconstructed as a sequence of plausible inferences which, for example, include the following very simple ones:

(10) Premises:
(a) [$0<|$If there are sentences conforming to the grammatical patterns of English, then there is a mental grammar. $|_{ST} <1$]
(b) $0<|$The sentence *Amy ate two peanuts* conforms to patterns of English.$|_{SJ} < 1$.
Conclusion:
(c) $0<|$There is a mental grammar.$_{I(10)}<1$

(11) Premises:
(a) [$0<|$If there are sentences which are nonsense, but conform to the grammatical patterns of English, then there is a mental grammar.$|_{ST} <1$]
(b) $0 < |$The sentence *Colorless green ideas sleep furiously* is nonsense and conforms to patterns of English.$|_{SJ} < 1$.
Conclusion:
(c) $0 < |$There is a mental grammar.$|_{I(11)} < 1$

András Kertész: The armchair in the laboratory
Argumentum 10 (2014), 359-371
Debreceni Egyetemi Kiadó

(12) Premises:
(a) [0<|If there are sentences in which not all words are real English words and which conform to the grammatical patterns of English, then there is a mental grammar. $|_{ST}$ <1]
(b) 0 < |In the sentence
Twas brillig, and the slithy toves
Did gyre and gimble in the wabe…
not all words are real English words and it conforms to patterns of English.$|_{SJ}$<1.
Conclusion:
(c) 0 < |There is a mental grammar.$|_{I(12)}$ < 1

The square brackets indicate that the statement within them is a latent background assumption. *ST* in the subscript refers to the source of the statement which is the theory Jackendoff accepts. In the conclusions the subscripts *I(10), I(11)* and *I(12)* stand for the inferences in (10), (11) and (12) as indirect sources of (10)(c), (11)(c) and (12)(c), respectively.

The relationship between (10)(c), (11)(c) and (12)(c) is characterised by the process of plausible argumentation in the course of which their plausibility values change. (10)(c) is, in the next subcycle of the argumentation process, retrospectively re-evaluated by (11)(c). The retrospective re-evaluation is prismatic, too, because through the premise (11)(b), new information is taken into consideration. As a result, the plausibility value of (11)(c) is higher than that of (10)(c). Analogously, (12) cyclically, prismatically and retrospectively re-evaluates (10)(c) and (11)(c) and thus leads to a plausibility value of (12)(c) higher than that of the former. Accordingly, (12)(c) is the provisional solution to the initial problem (7) in deciding between the alternatives (8)(a) and (b) by deeming (8)(b) implausible and assigning (8)(a) a higher plausibility value. In sum, the provisional result of the argumentation process can be summarised like this:

(13) (a) 0 < |It is not the case that there is no mental grammar.$|_{S}$ < 1
(b) 0 < |There is a mental grammar.$_{I(10)}$< |There is a mental grammar.$|_{I(11)}$ < |There is a mental grammar.$|_{I(12)}$ < 1

Now, after having seen how, during the plausible argumentation process, grammaticality judgments as the results of introspective thought experiments lead to the solution of the problem raised, we may turn to (P) and ask what kind of relationship there is between such introspective thought experiments and real experiments in Jackendoff's theory. In this respect, the following passage of Jackendoff's (1994: 48) argumentation is instructive:

(14) "Ideally, we might want to check these experiments out by asking large numbers of people under controlled circumstances, and so forth."

In (14), the expression "these experiments" refers to introspective thought experiments in Thomason's sense and thus represents the category mistake committed by Jackendoff already mentioned.

Furthermore, the quotation suggests that such thought experiments might be simply continued by real experiments (cf.: "asking large numbers of people under controlled circumstances"). Here the important insight is that in this way an introspective thought

experiment – provided that it has been carried out appropriately[5] – can be turned into a real one. But if this is so, then three consequences follow. First, according to Dąbrowska's finding

(15) "linguists' judgments are shown to diverge from those of nonlinguists. […] it is clear that linguists' judgments are not representative of the population as a whole, and hence syntacticians should not rely on their own intuitions when testing their theories". (Dąbrowska 2010: 1)

Consequently, if we accept the finding reported on in (15), then it is not the case that the linguist's thought experiment can be simply continued by asking many other people and thus turned into a real experiment.

Second, the p-model interprets the word "check" as the retrospective re-evaluation of the series of introspective thought experiments. Accordingly, in this interpretation, (14) describes what the p-model means by the dynamic process of plausible argumentation in the course of which previously assumed information is cyclically, prismatically and retrospectively re-evaluated.

The third consequence is that even if one accepts that asking many participants under controlled circumstances is a real experiment from the point of view of the researcher, each of the participants has to carry out his/her own thought experiment. Consequently, an alleged real experiment based on a questionnaire eliciting grammaticality judgments from the participants is nothing but a set of introspective thought experiments. In sum: what happens is merely that the armchair has been put into the laboratory, and the experimental linguist is a disguised armchair linguist.

In the next section we will discuss an example that illustrates the relationship between introspective thought experiments and real experiments from the perspective of the latter.

5 Second example: Introspective thought experiments as components of real experiments in cognitive linguistics

It is commonplace that in the classical contributions to conceptual metaphor research, claims about the existence of metaphorical concepts are based merely on the authors' intuitions about the interrelatedness of the meanings of certain expressions. Lakoff & Johnson (1980) take it for granted that the linguists' intuitions about the meanings of the expressions mirror conceptual structure.[6]

Now, if introspection is conceived of as thought experiment, and if conceptual metaphor research relies on introspection, then the sources that conceptual metaphor research makes use

[5] See Rákosi (2012) on the criteria that psycholinguistic experiments are expected to meet.

[6] Gibbs (2007: 3) summarises this stance in cognitive linguistics as follows:
"Despite their differences with generative linguists, cognitive linguists mostly employ traditional linguistic methods of examining native speakers' intuitions about the grammaticality and meaningfulness of linguistic expressions in order to uncover idealized speaker/hearer linguistic knowledge."
Itkonen writes:
"Cognitive Linguistics was born out of a feeling of dissatisfaction with generative linguistics. In the beginning, at least, this feeling did *not* extend to the type of *data* that the 'cognitive' linguist was supposed to account for. The data in Lakoff (1987), for instance, could be found in any generativist publication. […] the use of linguistic intuition plays exactly the same central (and even unique) role in cognitive linguistics as in generative linguistics." (Itkonen 2003: 69-70; emphasis as in the original)

of should be treated as thought experiments, too. In Lakoff and Johnson's book it is both introspective thought experiments and stage setting thought experiments in Thomason's sense that support the authors' main thesis:

(16) "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 & Johnson 1980: 3)

However, current conceptual metaphor research focuses on real experiments which seem to modify the methodology of the field substantially. During the past decade, a great number of real experiments have been conducted in order to extend and/or to check the research findings of classical approaches to conceptual metaphor research. Now, we will exemplify the relationship between the classical introspective thought experiments and real experiments by referring to Gibbs et al. (2004).

Gibbs et al. (2004) focus on the conceptual metaphor DESIRE IS HUNGER, and intend to show by real experiments that people's understanding of metaphorical expressions about human desires is motivated by their embodied experiences related to feeling hungry.[7] They carried out two experiments. In this section, we will reconstruct Gibbs et al.'s (2004) line of argumentation in a fragmentary and simplified manner.

(16), which had been obtained in part by introspective thought experiments, gave rise to a series of further tenets. As is well known, one of the tenets inferred from (16) in the literature on conceptual metaphors is the following (Gibbs et al. 2004: 1192):

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

In the literature, a complex process of plausible argumentation leads from (16) to (17). This complicated chain of inferences consisting of many argumentation cycles can be subsumed under the pattern of a plausible *modus ponens*:

(18) Premises:
(a) $0 < |\text{If (16), then (17)}|_{SCMR} < 1$
(b) $0 < |(16)|_{SCMR} < 1$
Conclusion:
(c) $0 < |(17)|_{I(18)} < 1$

Here *SCMR* stands for conceptual metaphor research as the direct source of the plausibility of the statements at issue.

Nevertheless, Gibbs et al. are not satisfied with (17) as discussed by Lakoff and Johnson and others. The plausibility value of (16) – from which (17) was inferred – is evaluated as considerably lower than it had been as the result of Lakoff and Johnson's (1980) thought experiments, i.e. it *decreases* at the start of the experimental report. The reason is that the authors consider the introspective data stemming from thought experiments to be *unreliable sources* in that they contrast them with the *reliability* of real experiments (Gibbs et al. 2004: 1207).

[7] For a detailed evaluation of this paper, see Csatár (2011).

András Kertész: The armchair in the laboratory
Argumentum 10 (2014), 359-371
Debreceni Egyetemi Kiadó

So, in order to decide whether (17) is plausible or not, they introduce the conceptual metaphor DESIRE IS HUNGER and put forward the following hypothesis:

(19) |The more prominent parts of the experimentees' hunger experiences are invariantly mapped onto their different concepts for desire. $|_{SCMR} = 0$

Here again, as in the previous example, this stage of the plausible argumentation process is informationally underdetermined because (19) is neither plausible nor implausible on the basis of the sources at the authors' disposal and is therefore, in the sense of the p-model, problematic. However, should (19) turn out to be plausible, then it could support (17) in the course of the plausible argumentation process through a series of intermediate argumentation steps that can be subsumed under the following pattern:

(20) Premises:
(a) $0 < |$If (19), then (17)$|_G < 1$
(b) $0 < |(19)|_{??} < 1$;
Conclusion:
(c) $0 < |(17)|_{I20} < 1$.

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 (19), and we have indicated this by the question marks.

Should the authors be able to assign a plausibility value to (19) by finding sources that they assume to be more reliable than Lakoff and Johnson's thought experiments, then with respect to (16), the structure of the inference under which the subsequent chain of further plausible inferences can be subsumed, would be that of *reduction*:[8]

(21) Premises:
(a) $0 < |$If (16), then (17)$|_{SCMR} < 1$
(b) $0 < |(17)|_{SCMR} < 1$
Conclusion:
(c) $0 < |(16)|_{I(21)} < 1$

Since (17) supports (16) (through the inference in (21)) and (17) is supported by (19) (see the inference in (20)) – whereby the source of (19) is missing – Gibbs et al.'s argumentation process is expected to find reliable sources that assign a plausibility value to (19).

During their argumentation in the experimental report, Gibbs et al. make use of three kinds of sources that jointly provide (19) with a plausibility value. The first is a methodological principle that says that "philosophical speculation is not enough", but "[e]mpirical research is needed to establish connections between embodiment and metaphor in thought and language" (Gibbs et al. 2004: 1207). Thereby, there is a latent background assumption according to which it is real experiments that make a particular research empirical.

Another source of the plausibility of (19) is two real experiments which investigate how American and Brazilian Portuguese experimentees' hunger experiences relate to their

[8] Using the notation of the p-model, the pattern of reduction is: $0<|$If *A*, then $B|_S<1$; $0<|B|_S<1$; therefore, $0<|A|_I<1$. Polya evaluates this pattern as "the simplest and most widespread pattern of plausible reasoning" (Polya 1948: 222). The term 'reduction' goes back to Łukasiewicz (1970 [1912]: 7).

conceptualization of desire. In the first experiment the experimentees (American and Brazilian students) were given a random list of symptoms that were assumed to result from somebody's being hungry. Their 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. The authors' conclusion was that there were significant regularities in people's embodied experiences of hunger. In the second experiment, the experimentees had to rate on a 7-point scale whether the expressions at issue were acceptable for them as ways of talking about desire in their mother tongue.

However, both real experiments included introspective thought experiments – the latter are the third source of the plausibility of (19). Thereby, the authors also committed the same category mistake that has been mentioned in the previous section. Gibbs et al. (2004) prepared the experiments by gathering a series of expressions such as *he hungers for recognition – he thirsts for recognition; he hungers for adventure – He thirsts for adventure*, etc. Obviously, the expressions were simply collected on the basis of the authors' native speaker knowledge of American English and Brazilian Portuguese – that is, *by introspection*, and it was postulated that there is a conceptual metaphor DESIRE IS HUNGER underlying them. In the first experiment, the experimenters compiled a list of symptoms that might be associated with hunger and they divided the list into three categories: local symptoms, general symptoms and behavioural symptoms. Each of these categories was further subdivided into closely related, possibly related and not related symptoms. The compilation of the list and its subdivisions were based on the experimenters' introspection. This also means that both the result of the linguistic analysis and the result of this experiment can be traced back to the same kind of data source, namely, introspection. When the experimentees had to rate the items according to the scale mentioned with respect to the question of whether they had experienced the effects listed when feeling hungry, the only source of their answers was introspection. In addition, the second real experiment was explicitly so designed that "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). Accordingly, in this way the real experiments conducted by Gibbs et al. consisted of a set of introspective thought experiments carried out by the participants.

Consequently, if introspection is a kind of thought experiment, then both real experiments included components that correspond to thought experiments: here again, the armchair is in the laboratory.

6 Conclusions

The application of the p-model to the above two examples yields the following solution to our problem (P) as raised in Section 1:

(i) If one interprets introspection as thought experiment and if introspection still assumes a central role in linguistics, then thought experiments are central to the latter.

(ii) Both examples witnessed that there is a multifaceted interaction between introspective thought experiments and real experiments in linguistics: thought experiments may be components of real experiments, real experiments are motivated by introspective thought experiments, and the results of real experiments retrospectively re-evaluate those of introspective thought experiments.

(iii) The application of the p-model has shown that this multifaceted interaction of introspective thought experiments and real experiments is governed by the dynamic process of plausible argumentation in the course of which pieces of information are cyclically, prismatically and retrospectively re-evaluated.

(iv) In both cases we have seen that the authors committed the category mistake of confusing real experiments with thought experiments. If something that is not a real experiment is claimed to be one and it is accompanied by the rhetoric of empiricism, then the methodology of linguistics permitting such a misconception is questionable.

(v) However, one of the reasons why this category mistake has been committed is that there is no avoiding the combination of real experiments and thought experiments. The sophisticated and reflected integration of real and thought experiments could contribute to the advancement of the field. The laboratory must be furnished with armchairs.

References

Brown, J.R. & Fehige, Y. (2011): Thought experiments. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/thought-experiment/. Accessed 12th July 2013.

Chomsky, N. (1986): *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Csatár, P. (2001): Principles of integrating psycholinguistic experiments in metaphor research. Parts I-II. *Sprachtheorie und germanistische Linguistik* 21.1, 3-24; 21.2, 109-132.

Dąbrowska, E. (2010): Naive vs. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27, 1-23.

Gibbs, R.W. Jr. (2007): Why cognitive linguists should care more about empirical methods. In: Gonzáles-Márquez, M., Mittelberg, I., Coulson, S. & Spivey, M.J. (eds.): *Methods in Cognitive Linguistics.* Amsterdam & Philadelphia: John Benjamins, 2-18.

Gibbs, R.W. Jr., Lima, P.L.C. & Francoso, E. (2004): Metaphor is grounded in embodied experience. *Journal of Pragmatics* 36, 1189-1210.

Haider, H. (2009): The thin line between facts and fiction. In: Featherston, S. & Winkler, S. (eds.): *The Fruits of Empirical Linguistics. Vol. 1: Process*. Berlin & New York: de Gruyter, 75-102.

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

Jackendoff, R. (1994): *Patterns in the Mind*. New York: Harper.

Kertész, A. (2014): The puzzle of thought experiments in Conceptual Metaphor Research. *Foundations of Science* (DOI: 10.1007/s10699-014-9357-z; in press).

Kertész, A. & Rákosi, Cs. (2009): Cyclic vs. circular argumentation in the Conceptual Metaphor Theory. *Cognitive Linguistics* 20, 703-732.

Kertész, A. & Rákosi, Cs. (2012): *Data and Evidence in Linguistics: A Plausible Argumentation Model*. Cambridge: Cambridge University Press.

Kertész, A. & Kiefer, F. (2013): From thought experiments to real experiments in pragmatics. In: Capone, A., Lo Piparo, F. & Carapezza, M. (eds.): *Perspectives on Philosophy and Pragmatics*. Berlin, Heidelberg & New York: Springer, 53-86.

Kertész, A. & Rákosi, Cs. (2014): Thought experiments and real experiments as converging data sources in pragmatics. In: Kertész, A. & Rákosi, Cs. (eds.): *The Evidential Basis of Linguistic Argumentation.* Amsterdam & Philadelphia: John Benjamins, 221-269.

Lakoff, G. (1987): *Women, Fire and Dangerous Things*. Chicago: Chicago University Press.

Lakoff, G. & Johnson, M. (1980): *Metaphors We Live By*. Chicago: University of Chicago Press.

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

Moue, A., Masavetas, K.A. & Karayianni, H. (2006): Tracing the developments of thought experiments in the philosophy of natural sciences. *Journal for General Philosophy of Science* 37, 61-75.

Polya, G. (1948): *How to Solve It*. Princeton: Princeton University Press.

Rákosi, Cs. (2012): The fabulous engine: strengths and flaws of psycholinguistic experiments. In: Kertész, A., Schwarz-Friesel, M. & Consten, M. (eds.): *Converging Data Sources in Cognitive Linguistics.* Special Issue of *Language Sciences* 34/6, 682-701.

Rescher, N. (1976): *Plausible Reasoning*. Assen & Amsterdam: Van Gorcum.

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

Schütze, C.T. (1996): *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: The University of Chicago Press.

Schütze, C.T. & Sprouse, J. (2013): Judgment data. In: Podesva, E.J. & Sharma, D. (eds.): *Research Methods in Linguistics*. Cambridge: Cambridge University Press, 27-50.

Schwarz-Friesel, M. (2012): On the status of external evidence in the theories of cognitive linguistics: compatibility problems or signs of stagnation in the field? Or: why do some linguists behave like Fodor's input systems? In: Kertész, A., Schwarz-Friesel, M. & Consten, M. (eds.): *Converging Data Sources in Cognitive Linguistics.* Special Issue of *Language Sciences* 34/6, 656-664.

Thomason, S.G. (1991): Thought experiments in linguistics. In: Horovitz, T. & Massey, G.J. (eds.): *Thought Experiments in Science and Philosophy*. Savage MD: Rowman and Littlefield, 247-257.

András Kertész
MTA-DE Research Group for Theoretical Linguistics
University of Debrecen
H-4010 Debrecen, Pf. 47
kertesz.andras@arts.unideb.hu