

In: Kertész, A., Rákosi, Cs. (eds.): *The Evidential Basis of Linguistic Argumentation*. Amsterdam & Philadelphia: Benjamins, 2014, 221-270.

Chapter 9

Thought experiments and real experiments as converging data sources in pragmatics

András Kertész
Csilla Rákosi

Abstract

The aim of the present chapter is to reconstruct the relationship between thought experiments and real experiments in pragmatics. In the first part, Searle's (1969) thought experiment on *Kennst du das Land, wo die Zitronen blühen?* is analysed with the help of the p-model. The second part reconstructs the argumentation structure of the real experiments reported on in Holtgraves and Ashley (2001), and shows that the latter and Searle's thought experiment closely interact. Basically, their relationship is characterised by the dynamic process of the cyclic, prismatic and retrospective re-evaluation of information in the sense of the p-model. However, they have different strengths and weaknesses. Accordingly, the generalised finding is that the conscious integration of real and thought experiments as data sources within some research may considerably enhance the reliability of the hypotheses.

1. Introduction

One of the most important insights obtained in the course of the current discussion on linguistic data and evidence is that in many cases it is not sufficient to restrict one's data to just one data type and data source. Rather, the present state of the discussion seems to suggest the integration of several data sources and data types. However, this desideratum is highly problematic. In particular, it is anything but clear which data sources and data types can be integrated, how integration can be achieved and under what conditions. With respect to these questions, the relationship between thought experiments and real experiments in pragmatics is especially challenging.

Austin's, Searle's, Grice's and others' thought experiments shaped the classical topics of pragmatics; moreover, without further thought experiments later approaches could not have been elaborated. Psycholinguistic experiments were sparse and did not gain ground for decades (see Noveck & Sperber 2007: 2). By now they have, however, become increasingly popular. Basically, this development could proceed in two directions.

The first possible direction is *the replacement of thought experiments by real ones*. The rapid development of experimental pragmatics has triggered decisive changes in the state of the art. The relevance of the current focus on real experiments in pragmatics is beyond doubt: the rise of experimental pragmatics has been accompanied by the introduction of new methods, the detection of new research topics, the publication of new findings and the restructuring of the scientific communities involved. Furthermore, there has also been a revision of earlier hypotheses and assumptions, solutions to a series of unsolved problems were proposed and attempts were made at substantial modifications of methodological rules applied in pragmatics research. For example, in Noveck and Sperber's characterisation, traditional pragmatics relied on pragmatic intuitions which must be regarded as results of thought experiments and cannot be interpreted as "facts" in the same sense as data which are normally regarded as objects of linguistic theorising:

It is a mistake to believe that pragmatic intuitions of the kind used in pragmatics are data of the same kind as semantic intuitions used in semantics. Genuine pragmatic intuitions are those that addressees have about the intended meaning of an utterance addressed to them. Quite generally, pragmatic intuitions invoked in theoretical pragmatics are not about actual utterances addressed to the reader of a pragmatic article, but about *hypothetical cases involving imaginary or generic interlocutors*. Pragmatic intuitions on hypothetical utterances have proved useful in a variety of ways, but it is important to keep in mind that these are not about how an utterance is interpreted, but about how an utterance **would be interpreted if** it were produced in a specific situation by a speaker addressing a listener, with referring expressions having actual referents, and so on. These intuitions are *educated guesses* – and, no doubt, generally good ones – *about hypothetical pragmatic facts*, but are not themselves pragmatic facts and they may well [be] in error. That is, we may be wrong about how, in fact, we would interpret a given utterance in a given context. (Noveck & Sperber 2007: 2f.; bold emphasis as in the original, italics added)

The features mentioned in this quotation – intuition, hypothetical cases, imaginary situation, educated guess – strongly resemble those features which the philosophical literature on thought experiments discusses. Although there are many different views of the nature of thought experiments, most authors agree that a thought experiment is an answer to a ‘What would be the case if?’ question. This answer is hypothetical: it makes use of an imaginary situation and intuition. Therefore, as a point of departure, the use of pragmatic intuition in the sense mentioned in the quotation basically seems to correspond to what is generally meant by thought experiments. If so, then it follows that neither pragmatic intuitions nor results gained from other kinds of thought experiments can be used for testing hypotheses, they can only motivate the raising of hypotheses. This is in accord with the requirements of the standard view of linguistic data and evidence that thought experiments (and particularly, pragmatic intuitions) cannot play the role of evidence for or against hypotheses.¹ From this Noveck and Sperber conclude that they have to be replaced by experimental data:

Besides helping compensate for the inherent limits of pragmatic intuitions, *an experimental approach can provide crucial evidence* when deciding between alternative theories that may agree on the content of the interpretations of utterances, but that have different implications regarding the cognitive mechanisms through which these interpretations are arrived at. (Noveck & Sperber 2007: 2f.; emphasis added)

According to this view, there is a kind of *linear development* in pragmatics as a result of which the method of thought experimentation should give way to experimental pragmatics.²

The second possibility is to acknowledge that experimental pragmatics has not been built on a ground which was prepared by thought experiments alone but that thought experiments and real experiments are intertwined and co-exist:

¹ For the notion of the standard view of the analytical philosophy of science see Kertész & Rákosi (this volume a, 2012: Section 3.3).

² In a considerably milder fashion, Noveck & Reboul (2008: 5f.) evaluate the appearance of experiments in pragmatics as the result of a linear development, too, where thought experiments are interpreted only as (historical) starting points. Nevertheless, they interpret the relationship between “traditional” pragmatics and experimental pragmatics not as a conflict or opposition but as an example of “how the cognitive sciences can work together in harmony” (Noveck & Reboul 2008: 5).

While the invention of (more or less fanciful) stories still plays a role in standard pragmatic argumentation, it appears that even in experimental pragmatics stories are an important part of methodology. (Meibauer 2012: 775)

At present, one cannot decide which of these alternative directions future development will take, and whether there are further possibilities as well. Therefore, one of the most relevant, but still unclarified, methodological aspects of current pragmatics is *the relationship between thought experiments and real experiments*. Thus, the present chapter centres on the problem (P) below:

(P) What kind of relationship is there between thought experiments and real experiments as data sources in pragmatics?

In order to get an insight into the nature of this relationship, we will apply Kertész & Rákosi's (2012) p-model as summarised in Kertész & Rákosi (this volume b).

In Section 2 we will provide a rough overview of the philosophical literature on thought experiments. We will highlight two points. First, according to several philosophers of science, there is a close relationship between real and thought experiments. Second, thought experiments are interpreted in a part of the current literature as a special kind of reasoning, and experimental narratives are supposed to be constitutive elements of both real and thought experiments. Therefore, the application of an argumentation theoretical model to real and thought experiments seems to be well-motivated. We have already proposed a model of psycholinguistic experiments based on the p-model (cf. Kertész & Rákosi 2012, Rákosi 2011a,b, 2012). Since the p-model characterises real experiments in such a way that it assumes that they are organised by a plausible argumentation process, we will try to characterise thought experiments along similar lines (see also Kertész 2010, Kertész & Kiefer 2013).

In order to illustrate how the p-model works and in order to exemplify how it yields a possible solution to (P), Section 3 will be devoted to a case study. We will reconstruct the argumentation structure of one of Searle's (1969) thought experiments.

Section 4 will focus on another case study: the analysis of real experiments as discussed in Holtgraves & Ashley (2001) which are related to the thought experiment discussed in Section 3.

On the basis of our analyses in Sections 3 and 4, in Section 5 we will reveal different aspects of the relationship between Searle's thought experiment and Holtgraves & Ashley's real experiments. After generalising our results, we will obtain the solution to (P).

Finally, in Section 6 we will summarise our findings and relate them to the contemporary discussion on data and evidence in linguistics.

2. On the state of the art

2.1. Current views on thought experiments in science

Thought experiments have been highly appreciated and frequently applied tools in science since Galileo:

There is widespread agreement that thought experiments play a central role both in philosophy and in the natural sciences and general acceptance of the importance and enormous influence and value of some of the well-known thought experiments in the natural sciences, like Maxwell's demon, Einstein's elevator or Schrödinger's cat. The 17th century saw some of its most brilliant practitioners in Galileo, Descartes, Newton, and Leibniz. And in our own time, the creation of quantum mechanics and relativity are almost unthinkable without the crucial role played by thought experiments. (Brown & Fehige 2011)

In contrast, in the philosophy of science – under the influence of the harsh criticism of Pierre Duhem and the standard view of the analytical philosophy of science – for a long time they were attributed only a psychological relevance as being solely a means of discovery but not of the justification of theories.³ Partly on the basis of the rediscovery of Ernst Mach's views on the subject, over the last two decades or so thought experiments have become an attractive topic in the philosophy of science and there has been a radical turn in the way they are evaluated. For most authors the focus no longer seems to be on deciding whether or not they are legitimate tools in scientific theorising but rather on metascientific reflection concerning their common features, classification, structure, functioning, limits, etc.

Although there is no widely accepted definition of the concept 'thought experiment', most authors agree that *argumentation* belongs to the constitutive elements of thought experiments. Nevertheless, there are several markedly different views about the nature and role of these argumentative aspects.

As we have already mentioned, many authors agree that thought experiments consider the consequences of a *hypothesised state of affairs*. However, this assumption gives rise to different lines of reasoning. For example, John Norton assumes that thought experiments "are just ordinary argumentation, disguised in some vivid picturesque or narrative form. As a result, they can do nothing more epistemically than ordinary argumentation." (Norton 2002: 1)

Norton's position did not find general acceptance and became the object of severe criticism. For example, Nersessian (1993) argues that although argumentation is one of their constitutive elements, thought experiments are basically of a non-argumentative character:

While I agree with Norton (1991) that thought experiments can often be reconstructed as arguments, the modeling function cannot be supplanted by an argument. As Norton acknowledges, the argument can be constructed only after the fact. That is, on my account, the argument is not evident until after the thought experiment has been constructed and executed. [...] real tworld experimental outcomes can be recast in argument form as well, but no one would argue that the experiment can be replaced by the argument. (Nersessian 1993: 297)

As this quotation suggests, Nersessian (1993: 292) distinguishes between the "original" thought experiment, on the one hand, and the "narrative" account of it, on the other. The former involves, first of all, the construction and manipulation of a *mental model of some imagined situation* by the thought experimenter. The latter makes it possible that members of the scientific community are capable of reconstructing the mental model and performing the same manipulations to it.

In the same vein, while emphasising the role of argumentation, Gooding points out that thought experiments cannot be reduced to deductive inferences as Norton requires. According to his view, thought experiments – similarly to real experiments – involve several different

³ For more on the history of the philosophy of thought experiments, see e.g., Nersessian (1993), Kühne (2005), Moue et al. (2004).

kinds of activities and are procedural instead of being simply propositional and static. According to Gooding (1993: 280), thought experiments and real experiments alike have to be regarded as *forms of experimental reasoning*. This means that they go beyond pure theorising, i.e., generalising and systematising our experiences about the world. Both are special cases of the activity of “experimenting”, i.e., making experiences by *conducting strictly regulated procedures under special, carefully controlled circumstances*. Similarly to Nersessian, Gooding (1993) differentiates between the process of real/thought experimentation and its narrative account:

It is still the case that most people – including most scientists – encounter experiments through narratives. The narrative elements of arguments that invoke real experiments require the same ability to participate. [...] To explain the force of an experiment it helps to understand it as a process to be worked through, rather than as a logical structure. (Gooding 1993: 283)

Gooding remarks that his and Nersessian’s standpoints are close to each other. They share a common general view according to which thought experiments and real experiments are cognates. Nevertheless, they take different perspectives: while Gooding’s point of reference seems to be real experiments in physics, Nersessian’s position can be interpreted as that of a cognitive psychologist.⁴

2.2. *Points of departure for the solution of (P)*

We accept Nersessian’s and Gooding’s view that real and thought experiments are kindred. Nevertheless, we see the role of argumentation in both kinds of experiments differently from these two authors. In Kertész & Rákosi (2012), Rákosi (2011a,b) and Rákosi (2012) we presented a model of psycholinguistic experiments and argued that argumentation has a *twofold task* in real experiments. First, the experimental process is organised and conducted by an argumentation process which remains the private affair of the experimenters. It aims at the clarification of the relationship among different components and stages of the experimental procedure in order to obtain a coherent and comprehensive picture of the state of affairs at issue from a multitude of uncertain, fallible and conflicting pieces of information. Second, the main steps and results of the experimental process are transformed to a piece of public argumentation as presented in the *experimental report*. In analogy to the term ‘experimental report’ used with respect to real experiments, in the case of thought experiments we will call such a presentation a *thought experimental report*. Both kinds of reports present those facets of the real/thought experimental process that might be *relevant* to the evaluation and acceptance of the results. To this end, relevant pieces of information are selected from the original, non-public argumentation and arranged into a well-structured chain of arguments leading from the problems raised to the data which have been obtained by carrying out the real/thought experiment. Nevertheless, the real/thought experimental report still contains traces of the original argumentation process organising the real/thought experimental process.⁵

⁴ “Like Nersessian (1991, 1993) I regard thought experiments as process-narratives to which visualization is essential. Where she has developed a mental-models account of the process, I emphasize the role of embodiment in visualization.” (Gooding 1993: 281)

⁵ Similarly to real experiments, thought experiments are not directly accessible to the reader but they can be *partially* reconstructed from the paper describing and summarising them. Therefore, the primary object of an argumentation theoretical reconstruction can be the paper presenting the results, the ‘(thought) experimental report’.

This chain of thought motivates the elaboration of a metatheoretical model of thought experiments which is based on the idea that the similarities of real and thought experiments as well as the peculiarities of the latter might be grasped from an argumentation theoretical point of view. We hypothesise that the p-model meets this desideratum. In the next subsection we will show how it can be applied to one of the most famous classical thought experiments in pragmatics: Searle's *Kennst du das Land, wo die Zitronen blühen?*⁶ With this thought experiment Searle argues both against Grice's explication of the notion of 'non-natural meaning' and for his own account of illocutionary acts. In accordance with the p-model, we will reconstruct the presentation of this thought experiment in Searle (1969) – that is, his thought experimental report – as a cyclic and prismatic process of the retrospective re-evaluation of the information at hand.

3. Case study 1: Thought experimental report in Searle (1969)

3.1. Setting up the starting p-context

The initial problem Searle raises is this:

- (1) “[W]hat is it for one to mean something by what one says [...]”? (Searle 1969: 42)

The *starting p-context* includes Grice's (1989a [1957]) explication of the notion of 'non-natural meaning' which is one of the central theses of his model of communication:⁷

- (2) “[t]o say that a speaker *S* meant something by *X* is to say that *S* intended the utterance of *X* to produce some effect in a hearer *H* by means of the recognition of this intention.” (Searle 1969: 43)

(2) receives a positive plausibility value on the basis of Grice (1989a [1957]) as a direct source. This source is part of the p-context of Searle's argumentation, but its reliability is not sufficient to assign to it a higher plausibility value. Therefore, Searle seems to be ready to acknowledge Grice's paper as a reliable source only to a limited extent and only provisionally. The reason for this decision is the circumstance that Searle does not find Grice's argumentation in support of (2) convincing, because, according to him,

- (3) (2) “fails to account for the extent to which meaning can be a matter of rules or conventions”. (Searle 1969: 43)

Should one accept (3), then (2) would be implausible; nevertheless, (3) is at this point not reliable enough to legitimise the rejection of (2). Thus, the starting p-context is *p-inconsistent* in the sense of Subsection 2.5 of Kertész & Rákosi (this volume b). Accordingly, Searle's aim can be characterised in such a way that he tries to increase the plausibility value of (3) and to decrease that of (2).

⁶ In Goethe's original poem, one finds *blüh'n* instead of *blühen*. In quoting Searle, we will preserve his way of quoting Goethe.

⁷ Following Searle, we will speak simply of 'meaning', because we are not concerned with such uses of 'mean' as e.g. *Clouds mean rain* which Grice (1989a [1957]) distinguishes from 'non-natural meaning'.

3.2. The elaboration of the starting p-context

First, Searle applies the *Exclusive Strategy* (Kertész & Rákosi, this volume b: Subsection 2.6.2), that is, he develops a p-context version based on (2) in order to examine whether it is a tenable alternative. Searle *extends* it by the consideration of a *new source* – namely, his *thought experiment*. For convenience we quote Searle’s thought experimental report at length:⁸

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

⁸ We will divide the quotations into smaller units in order to make it easier to refer to parts of the quotation to be analysed. The units do not necessarily correspond to paragraphs in the original text.

- i. We **must** [...] reformulate the Gricean account of meaning in such a way as to make it clear that one's meaning something when one utters a sentence is more than just randomly related to what the sentence means in the language one is speaking. In our analysis of illocutionary acts, we **must** capture both the intentional and the conventional aspects and especially the relationship between them." (Searle 1969: 44f.; italics as in the original, bold emphasis added)

(4)(a) delineates an imaginary situation. In (4)(b)-(c), plausible inferences are presented which provide an interpretation of some facets of the given situation with the help of the terminology of speech act theory. These inferences rely on the following data introduced into the p-context through the imaginary situation in (4)(a):

- (5) a. $0 < |$ In the situation depicted in (4)(a), John Searle says, *Kennst du das Land...* etc. because he wishes to get Italian soldiers to believe that he is a German soldier. $|_S < 1$
 b. $0 < |$ In the situation depicted in (4)(a), Italian soldiers do not know what the utterance *Kennst du das Land...* etc. means. $|_S < 1$
 c. $0 < |$ In the situation depicted in (4)(a), Italian soldiers can realise that the utterance *Kennst du das Land...* etc. is a German sentence. $|_S < 1$

The direct source of the plausibility of the statements in (5) is a compound of subsources, abbreviated as *S*. It is originated in Searle's everyday (linguistic) intuition. Nevertheless, we have no direct access to it but we can reconstruct it from the text of Searle (1969) by making use of our own (linguistic) intuition. These statements do not contain theoretical terms but everyday concepts.

Based on these data, in (4)(b) the characterisation of Searle's behaviour as described in (4)(a) is reformulated, that is, retrospectively re-evaluated in accordance with (2) in Gricean terms. Thereby, 'wish to get *X* to believe that *Y*' as used in (5)(a) is explicated as 'intending that *X* thinks that what was told him is *Y*' in the sense of (2):

- (6) $0 < |$ In the situation depicted in (4)(a), John Searle says, *Kennst du das Land...* etc. because he wishes to get Italian soldiers to believe that he is a German soldier. $|_S < 1$
 (= (5)(a))
 $[0 < |$ If in the situation depicted in (4)(a), John Searle says, *Kennst du das Land...* etc. because he wishes to get Italian soldiers to believe that he is a German soldier, then he *intends* that the German soldiers should think that what he is trying to tell them is 'I am a German soldier' when saying *Kennst du das Land...* etc. $|_S < 1]$
 $0 < |$ In the situation depicted in (4)(a), Searle *intends* that the German soldiers should think that what he is trying to tell them is 'I am a German soldier' when saying *Kennst du das Land...* etc. $|_{I(6)} < 1$ ⁹

Since the conclusion of (6) will play an important role in Searle's argumentation, we will represent it as hypothesis (7):

- (7) $0 < |$ In the situation depicted in (4)(a), Searle *intends* that the German soldiers should think that what he is trying to tell them is 'I am a German soldier' when saying *Kennst du das Land...* etc. $|_{I(6)} < 1$

⁹ In the conclusion of (6), *I(6)* stands for the inference in (6) as an indirect source of the conclusion. */p/* indicates the plausibility value of statement *p*; see Kertész & Rákosi (this volume b: Section 2.2, Example 3; Section 2.3).

The question in the last sentence of (4)(b) rests on the assumption that the application of (2) to the situation described in (4)(a) would lead to the conclusion that the meaning of Searle's words has to be identified with the content of his intention:

- (8) $0 < |$ In the situation depicted in (4)(a), Searle *intends* that the German soldiers should think that what he is trying to tell them is 'I am a German soldier' when saying *Kennst du das Land...* etc. $|_{I(6)} < 1$ (= (7))
 $0 < |$ The *meaning* of an utterance has to be identified with the speaker's *intention* related to this utterance. $|_G < 1$ (= (2))¹⁰

 $0 < |$ In the situation depicted in (4)(a), Searle *means* with his words 'I am a German soldier' when saying *Kennst du das Land...* etc. $|_{I(8)} < 1$

We will refer to the conclusion of inference (8) as hypothesis (9):

- (9) $0 < |$ In the situation depicted in (4)(a), Searle *means* with his words 'I am a German soldier' when saying *Kennst du das Land...* etc. $|_{I(8)} < 1$

This means that the datum in (5)(a) provides *weak evidence* for (9): the plausible inferences (6) and (8) establish a connection between (5)(a) and this hypothesis. Thus, the analysis of the imagined situation in (4)(a) based on Grice's theory of meaning, and especially, on hypothesis (2) leads to the conclusion that the meaning of Searle's words in the situation in (4)(a) has to be identified with the speaker's intention, that is, with the information that he wants to communicate. The current state of the p-context is as in Figure 1.¹¹

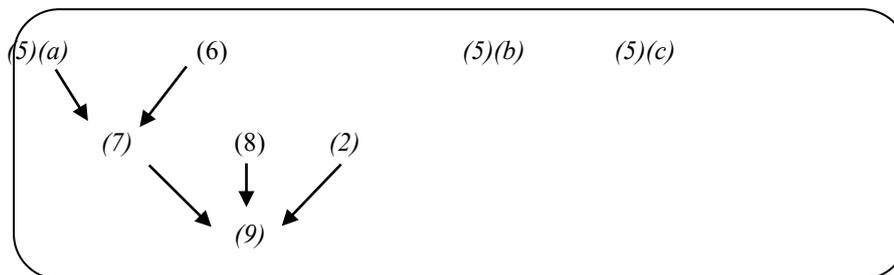


Figure 1: The state of the p-context after its extension with (4)(a)-(b)

(9) is, however, questioned immediately in the first sentence of (4)(c). The starting point of Searle's argumentation against (2) is the following datum:

- (10) $0 < |$ According to the rules of German as well as what Searle remembers, the words *Kennst du das Land...* etc. *mean* that 'Knowest thou the land where the lemon trees bloom?'. $|_S < 1$

The extension of the p-context with (10) makes it possible to draw the following plausible inference:

¹⁰ *G* indicates Grice (1989a [1957]) as a direct source.

¹¹ In Figures 1-10, hypotheses are in italics, and inferences are in normal script. Conclusions of inferences are connected with their premises and the given inference by simple arrows. Double dotted arrows indicate p-inconsistency.

- (11) $0 < |$ According to the rules of German as well as what Searle remembers, the words *Kennst du das Land...* etc. *mean* ‘Knowest thou the land where the lemon trees bloom?’. $|_S < 1$ (= (10))
 $[0 < |$ There is no difference between the *meaning* of these words according to the rules and conventions of the German language as Searle remembers and what Searle *means* with them in the given situation. $|_S < 1]$

 $0 < |$ In the situation depicted in (4)(a), Searle *means* with his words ‘Knowest thou the land where the lemon trees bloom?’. $|_{I(11)} < 1$

In (11), the latent background assumption identifies what Searle means by his utterance in this situation with the conventional meaning of his words. Thus, the argumentation process has led to the following hypothesis:

- (12) $0 < |$ In the situation depicted in (4)(a), Searle *means* with his words ‘Knowest thou the land where the lemon trees bloom?’. $|_{I(11)} < 1$

(8) and (11) lead to contradicting conclusions – clearly due to the circumstance that they use ‘mean’ in two different senses: in (8) in accordance with (2), that is, by identifying meaning with the speaker’s intention, and in (11) by identifying it with the conventional meaning of the uttered words. Figure 2 shows the current state of the p-context.

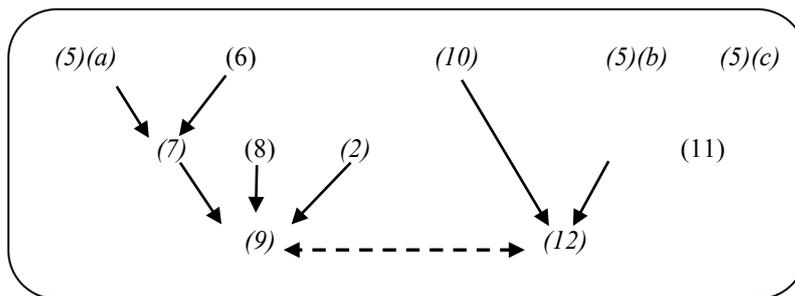


Figure 2: The state of the p-context after its extension with the first part of (4)(c)

Searle deems (12) more plausible than (9) and solves this contradiction in such a way that he gives up (9) and retains (12). Thus, the following indirect source presents itself:

- (13) $0 < |$ In the situation depicted in (4)(a), Searle *means* with his words ‘Knowest thou the land where the lemon trees bloom?’. $|_{I(11)} < 1$ (= (12))
 $0 < |$ In the situation depicted in (4)(a), Searle *intends* that the German soldiers should think that what he is trying to tell them is ‘I am a German soldier’ when saying *Kennst du das Land...* etc. $|_{I(6)} < 1$ (= (7))
 $[0 < |$ If in the situation depicted in (4)(a), Searle *intends* that the German soldiers should think that what he is trying to tell them is ‘I am a German soldier’ when saying *Kennst du das Land...* etc., but he *means* with his words ‘Knowest thou the land where the lemon trees bloom?’, then the *meaning* of this utterance cannot to be identified with the speaker’s *intention* related to this utterance. $|_S < 1]$

 $0 < |$ In the situation depicted in (4)(a), the *meaning* of Searle’ utterance cannot be identified with the speaker’s *intention* related to this utterance. $|_{I(13)} < 1$

Thus, the interpretation of the situation in (4)(a) comes to the conclusion that

(14) $0 < |$ In the situation depicted in (4)(a), the *meaning* of Searle's utterance cannot be identified with the speaker's *intention* related to this utterance. $|_{I(13)} < 1$

In the last sentence of (4)(c), Searle provides a possible analysis of the mechanism of the deceit applied in the story. There he seems to exploit the given situation with the help of the data in (5)(b) and (c):

(15) $0 < |$ In the situation depicted in (4)(a), the *meaning* of Searle's utterance cannot be identified with the speaker's *intention* related to this utterance. $|_{I(13)} < 1$ (= (14))

[$0 < |$ If in the situation depicted in (4)(a), the *meaning* of Searle's utterance cannot be identified with the speaker's *intention* related to this utterance, Italian soldiers do not know what the utterance *Kennst du das Land...* etc. means and they can realise that the utterance *Kennst du das Land...* etc. is a German sentence, then they will erroneously think that the sentence *Kennst du das Land...* etc. means 'I am a German soldier' according to the conventions of the German language. $|_S < 1$]

$0 < |$ If in the situation depicted in (4)(a), Italian soldiers erroneously think that the sentence *Kennst du das Land...* etc. means 'I am a German soldier' according to the conventions of German, then they should erroneously think that with the utterance of the sentence *Kennst du das Land...* etc. Searle *means* 'I am a German soldier'. $|_S < 1$

$0 < |$ In the situation depicted in (4)(a), Italian soldiers do not know what the utterance *Kennst du das Land...* etc. means. $|_S < 1$ (= (5)(b))

$0 < |$ In the situation depicted in (4)(a) Italian soldiers can realise that the utterance *Kennst du das Land...* etc. is a German sentence. $|_S < 1$ (= (5)(c))

$0 < |$ In the situation depicted in (4)(a), Italian soldiers may erroneously think that with the utterance of the sentence *Kennst du das Land...* etc. Searle *means* 'I am a German soldier'. $|_{I(15)} < 1$

As a result, a hypothesis was obtained that would be p-inconsistent with (9) that has already been rejected, but is p-consistent with (12):

(16) $0 < |$ In the situation depicted in (4)(a), Italian soldiers may erroneously think that with the utterance of the sentence *Kennst du das Land...* etc. Searle *means* 'I am a German soldier'. $|_{I(15)} < 1$

After the *coordination* of the p-context, the p-context is as in Figure 3.

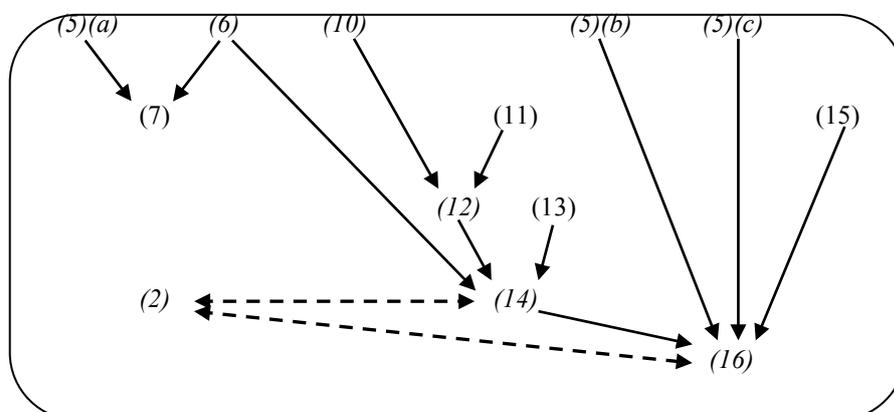


Figure 3: The state of the p-context after its extension with the second part of (4)(c)

As Figure 3 illustrates, the *Exclusive Strategy* applied by Searle in order to examine the p-context which can be built around hypothesis (2) failed. The reason is that the plausible

inferences drawn on the basis of the data available made hypotheses plausible which are incompatible with (2).

3.2.1. Subcycle 1: Analysis of Wittgenstein's thought experiment

Although Searle gave up (9), he does not *modify* the p-context by also deleting (2) from the p-context. The reason for this decision might be that at this point of his argumentation, he is not in the possession of an alternative hypothesis to (2). Therefore, he starts a new *subcycle* in order to formulate and test a rival hypothesis and to find further data for a well-founded solution to the starting problem. Nevertheless, it is important to remark that Searle's argumentation is enthymematic here. Of course, this makes the reconstruction of his argumentation more difficult and disputable.¹²

In (4)(d), one of Wittgenstein's thought experiments is referred to. The p-context is *extended* by a new source which is a quotation from § 510 of Wittgenstein's *Philosophical Investigations*.¹³ The starting point of Searle's argumentation is the following datum:

(17) $0 < | \text{Without further stage setting, when one says, } \textit{It's cold here}, \text{ one cannot mean 'It is warm here'} .|_S < 1$

At this point, Searle does not clarify what "further stage setting" means; only later does the reader receive explanation, in (4)(h). Here again, the reliability of this thought experiment depends crucially on the question of whether the reader is ready to accept this datum by judging it with the help of his/her own (linguistic) intuition. Based on (17), the following plausible inference presents itself:

(18) $0 < | \text{Without further stage setting, when one says, } \textit{It's cold here}, \text{ one cannot mean 'It is warm here'} .|_S < 1$ (= (17))
 $0 < | \text{If without further stage setting, when one says, } \textit{It's cold here}, \text{ one cannot mean 'It is warm here'}, \text{ then there are cases where, what one can mean is a function of what one is saying} .|_S < 1$

 $0 < | \text{There are cases in which, what one can mean is a function of what one is saying} .|_{I(18)} < 1$

(18) is based on a possible explanation of the datum in (17). A decisive element in the evaluation of Searle's argumentation is whether one is ready to accept the second premise, and via this, the conclusion of (18), which, for future reference, we highlight as (19):

(19) $0 < | \text{There are cases in which, what one can mean is a function of what one is saying} .|_{I(18)} < 1$

(19) is clearly p-consistent with (14), which states that the meaning of Searle's utterance in the situation depicted in (4)(a) cannot be identified with the speaker's intention related to this utterance.

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

¹³ The whole passage is this: "510. Make the following **experiment**: say 'It's cold here' and mean 'It's warm here'. Can you do it? – And what are you doing as you do it? And is there only one way of doing it?" (Wittgenstein 1958 [1953]: 140; italics as in the original, bold emphasis added)

(4)(f) continues the analysis of the Wittgensteinean situation and puts forward a plausible inference based on (19) in order to elucidate and generalise it. Searle's argument can be reconstructed in the following way:

- (20) $0 < |$ There are cases in which what one can *mean* is a function of what one is saying. $|_{I(18)} < 1 (= (19))$
 $[0 < |$ If there are cases in which what one can *mean* is a function of what we are saying, then *meaning* is more than a matter of *intention*, it is also at least sometimes a matter of *convention*. $|_S < 1]$

 $0 < |$ Meaning is more than a matter of *intention*, it is also at least sometimes a matter of *convention*. $|_{I(20)} < 1$

The analysis of the state of affairs referred to in Wittgenstein's thought experiment leads to the hypothesis that the meaning of an utterance cannot be identified with the speaker's intention in these cases but involves the conventional meaning of the sentence at issue as well. Thus, *a highly important result of Searle's thought experiment provided in the conclusion of inference (20) is a hypothesis which is a rival to (2):*

- (21) $0 < |$ Meaning is more than a matter of *intention*, it is also at least sometimes a matter of *convention*. $|_{I(20)} < 1$

At this point, Searle's argumentation turns back to the point where it had been suspended before the subcycle was started, and the new information obtained is added to the p-context. The state of the p-context as depicted in Figure 3 has been *retrospectively re-evaluated*. Data originating from Wittgenstein's thought experiment contributed to the increase of the plausibility value of $\sim(2)$, since besides (14) and (16), (21) is also p-inconsistent with (2). Thereby, (21) is a reformulated and re-evaluated version of $\sim(2)$. Since (19) covers (14) as a special case, it would be possible to construct a plausible inference with (14) as one of its premises and (21) as its conclusion. After its *coordination*, the p-context is as in Figure 4.

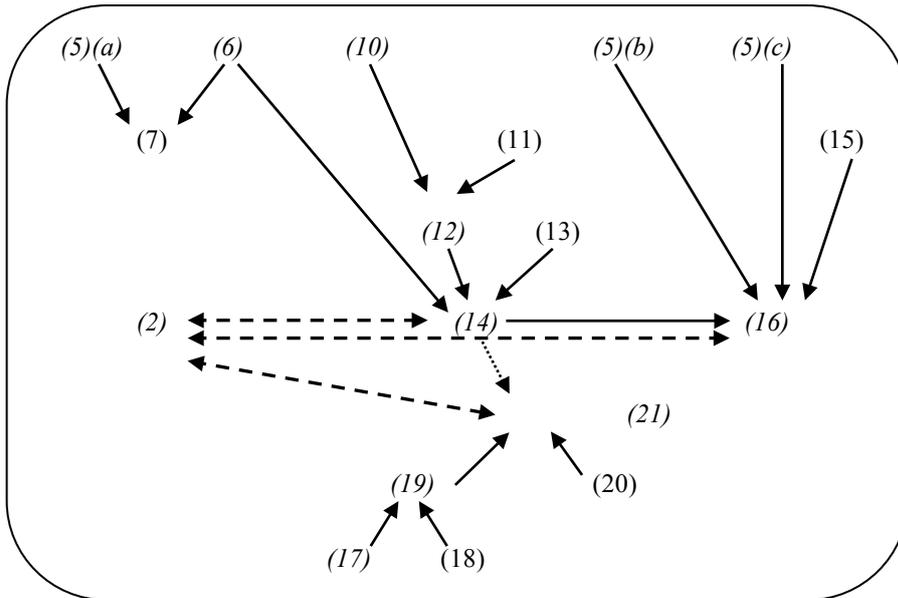


Figure 4: The state of the p-context after its extension with (4)(d)-(f)

3.2.2. *The continuation of the argumentation cycle: Further elaboration of a rival hypothesis*

After the coordination of the p-context, Searle turns to the application of the *Contrastive Strategy* in order to arrive at a decision between (2) and its rival, (21). In (4)(g), he *generalises* the conclusion which emerges from the application of the Gricean hypothesis in (2) to the imaginary situations delineated in (4)(a) and (4)(d):

- (22) $0 < |$ If the *meaning* of an utterance is identified with the speaker's *intention* related to this utterance, then any sentence can be uttered with any *meaning* whatever, given that the circumstances make possible the appropriate *intentions*. $|_S < 1$
 $0 < |$ In the situation depicted in (4)(a), the *meaning* of Searle's utterance cannot be identified with the speaker's *intention* related to this utterance. $|_{I(13)} < 1 (= (14))$
 $0 < |$ In the situation depicted in (4)(d), without further stage setting, when one says, *It's cold here*, one cannot *mean* that 'It is warm here'. $|_S < 1 (= (17))$

 $0 < |$ The *meaning* of an utterance cannot be identified with the speaker's *intention* related to this utterance. $|_{I(22)} < 1 (= \sim(2))$

That is, the plausible inference presented in (22) makes (2) implausible, because it shows that (2) overgeneralises. The p-context is as in Figure 5.

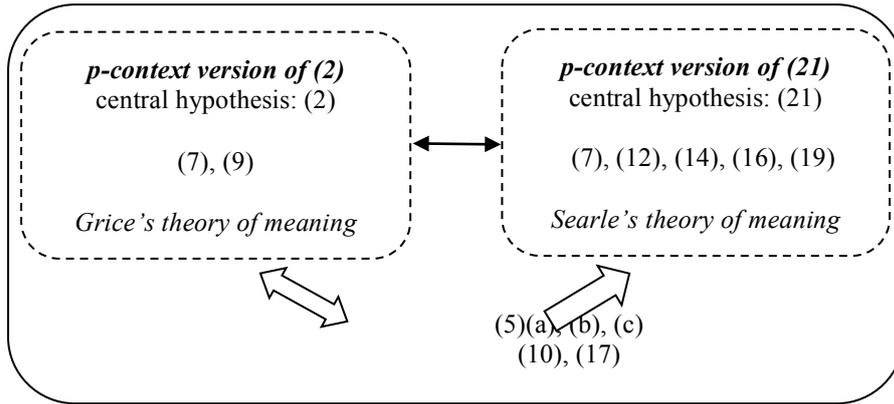


Figure 5: The p-context resulting from the application of (2) to the imaginary situations delineated in (4)(a) and (4)(d)

While the data stemming from the two thought experiments provide weak evidence for (21), they also provide weak evidence against (2). Therefore, since the p-context version of (2) is burdened with a high amount of p-inconsistency, Searle gives it up and preserves only its rival and elaborates further on it. He turns back to the Exclusive Strategy in order to find out whether the p-context version of (21) provides a satisfactory solution of the starting p-problem.

(17), (19), and (21) do not exclude the possibility that sometimes a situation can be established in which what one can mean is independent of what one is saying. As we have remarked with respect to (17) in Subsection 3.2.1, Searle does not provide any explanation of what “further case setting” means. Thus, it remains unclarified how cases in which utterance meaning can be identical with a completely different speaker’s intention can be distinguished from cases in which this is not possible. (4)(h) deals with this question and identifies the factors which are supposed to constitute the difference between the two types of cases. Since it is not clear from Searle’s formulation which of the two thought experiments – i.e. Searle’s or Wittgenstein’s – is referred to in this passage, our reconstruction will pertain to both of them:

- (23) $0 < |$ In the situation depicted in (4)(a), Searle *intends* that the hearer should think that what he is trying to tell him is ‘I am a German soldier’ when saying, *Kennst du das Land... etc.* $|_{(7)(6)} < 1$ (= (7))
 $0 < |$ According to the rules of German as well as what Searle remembers, the words *Kennst du das Land... etc.* mean ‘Knowest thou the land where the lemon trees bloom?’ $|_S < 1$ (= (10))
 $[0 < |$ If according to the rules of German, the words *Kennst du das Land... etc.* mean ‘Knowest thou the land where the lemon trees bloom?’, then these words are a device which is conventionally used as a means of producing quite different illocutionary effects than ‘I am a German soldier’ $|_S < 1]$
-
- $0 < |$ In the situation depicted in (4)(a), Searle *intends* to produce a certain effect by means of getting the hearer’s recognition of his intention to produce that effect, but the device he uses to produce this effect is one which is conventionally used as a means of producing quite different illocutionary effects. $|_{I(23)} < 1$
- (24) $0 < |$ In the situation depicted in (4)(d), one *intends* that the hearer should think that what one is trying to tell him is ‘It is warm here’ when saying, *It’s cold here.* $|_S < 1$
 $[0 < |$ According to the rules of English, the words *It’s cold here* mean ‘It’s cold here’ $|_S < 1]$

[0 < |If according to the rules of English, the words *It's cold here* mean 'It's cold here', then these words are a device which is conventionally used as a means of producing quite different illocutionary effects than 'It is warm here'.|_S < 1]

0 < |In the situation depicted in (4)(d), one *intends* to produce a certain effect by means of getting the hearer's recognition of his intention to produce that effect, but the device he uses to produce this effect is one which is conventionally used as a means of producing quite different illocutionary effects.|_{I(24)} < 1

As a next step, in (4)(h), relying, among other things, on the conclusion of (23) and (24), Searle generalises the explanation which could be applied to these two cases with the help of the following inductive inference:

(25) 0 < |In the situation depicted in (4)(a), Searle *intends* to produce a certain effect by means of getting the hearer's recognition of his intention to produce that effect, but the device he uses to produce this effect is one which is conventionally used as a means of producing quite different illocutionary effects.|_{I(23)} < 1

0 < |In the situation depicted in (4)(d), one *intends* to produce a certain effect by means of getting the hearer's recognition of his intention to produce that effect, but the device he uses to produce this effect is one which is conventionally used as a means of producing quite different illocutionary effects.|_{I(24)} < 1

0 < |In the situation depicted in (4)(a) and (4)(d), the stage setting or conditions which would permit one to say one thing and *mean* something totally unrelated are not present.|_S < 1

0 < |In the situation depicted in (4)(a) and (4)(d), the sentences mentioned cannot be uttered with a *meaning* identical to the speaker's *intention*.|_{(14) & (17)} < 1

[0 < |In every other case, too, if the circumstances do not make possible the appropriate *intentions* or (i) the sentence cannot be conventionally, by the rules governing the use of this sentence, used as a means of producing this illocutionary effect, and (ii) the stage setting or conditions which would permit one to say one thing and *mean* something totally unrelated are not present, then this sentence cannot be uttered with a *meaning* identical to the speaker's *intention*.|_S < 1]

0 < |A sentence can be uttered with a *meaning* only if the circumstances make possible the appropriate *intentions* and (i) the sentence can be conventionally, by the rules governing the use of this sentence, used as a means of producing this illocutionary effect, or (ii) the stage setting or conditions which would permit one to say one thing and *mean* something totally unrelated are present.|_{I(25)} < 1

From this it seems that (4)(a) and (4)(d) refer to one's linguistic behaviour under normal circumstances, that is, to a situation in which the speaker and the hearer know the language used well enough so that he/she cannot be mistaken or deceived about the conventional meaning of the words used, and both the speaker and the hearer are supposed to use language in a normal way, without any special modification of the meaning of words.

The conclusion of (25) is a retrospectively re-evaluated version of (21), that is, a further refinement of ~(2):

(26) A sentence can be uttered with a *meaning* if the circumstances make possible the appropriate *intentions* and (i) the sentence can be conventionally, by the rules governing the use of this sentence, used as a means of producing this illocutionary effect, or (ii) the stage setting or conditions which would permit one to say one thing and *mean* something totally unrelated are present.

3.3. *The comparison of the solutions to the problem raised in (1)*

On the basis of the above considerations we can state that a better solution can be built around (26) rather around (2), since the latter does not receive any support from other direct or indirect sources in this p-context and its presence in the p-context leads to p-inconsistencies.

3.4. *The final p-context*

Thus, the p-context is modified: (2) is given up and replaced by a rival hypothesis proposed by Searle, because only the latter is p-consistent with the data obtained in the thought experiment. Therefore, the *final p-context* does not contain (2) because it has been rejected and replaced by (21) and further elaborated as (26). The reconstruction of the inference in (4)(i) which presents the methodological consequences emerging from the two thought experiments analysed by Searle is as follows:

- (27) $0 < | \text{Meaning is more than a matter of } \textit{intention}, \text{ it is also at least sometimes a matter of } \textit{convention}. |_{I(20)} < 1$ (= (21))
 $0 < | \text{If } \textit{meaning} \text{ is more than a matter of } \textit{intention}, \text{ and it is also at least sometimes a matter of } \textit{convention}, \text{ then the analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them.} |_S < 1$

 $0 < | \text{The analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them.} |_{I(27)} < 1$

That is, the results of the two thought experiments motivate the following methodological principle:

- (28) The analysis of illocutionary acts must capture both intentional and conventional aspects of meaning and the relationship between them.

Figure 6 shows the final p-context.

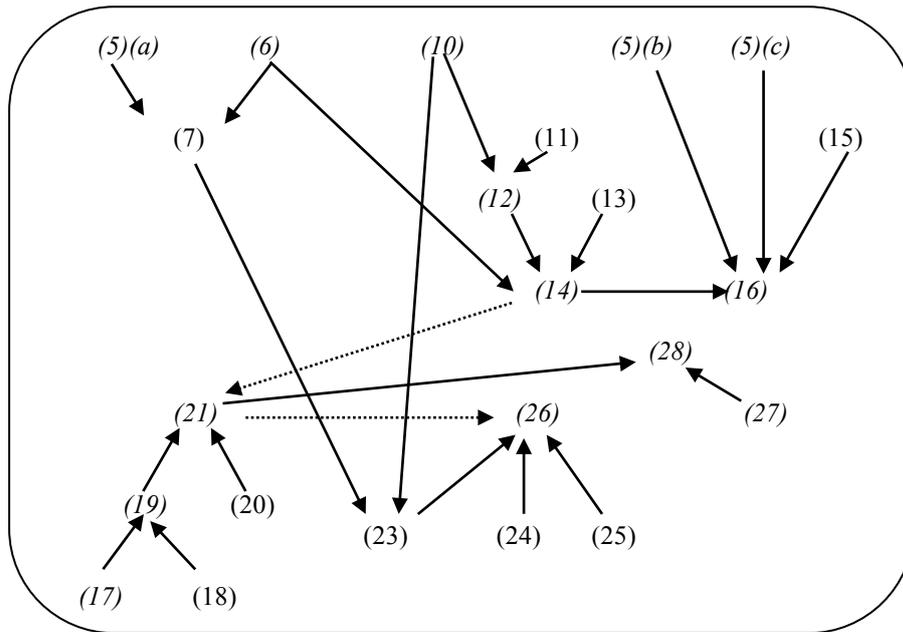


Figure 6: The final p-context

This solution to (1) is provisional and does not count as a resolution in the sense of Subsection 2.5 in Kertész & Rákosi (this volume b). It is exposed to further revision for at least three reasons.

First, the argumentation process continues with Searle’s (1969: 43) consideration of the problem of “what is it for something to have a meaning in analogy to the process we have just exemplified”. This problem is discussed on pp. 49-50 by contrasting Grice’s solution with Searle’s. The final p-context of this continuation includes the retrospective re-evaluation of (21)/(26) in the light of Searle’s solution to the above problem. We will not reconstruct this process here, because it is analogous to the one we have just analysed.

Second, as is well known, Searle’s account of meaning supported by the thought experiment in (4) turned out to be problematic in several respects again. Grice (1989b), Meggle & Ulkan (1979) and other contributions argued against (21)/(26); i.e. these authors re-evaluated the results of the plausible argumentation which we have reconstructed above.

Third, the counter-arguments against (21)/(26) motivated, among others, *real experiments*. In the next section we will analyse an example which illustrates how (21)/(26) can be retrospectively re-evaluated by real experiments.

We choose Holtgraves & Ashley (2001) as the subject matter of the case study that scrutinises a problem with the help of a real experiment which is directly connected to the thought experiment presented by Searle and analysed in the previous section. For lack of space, we will dispense with the reconstruction of Holtgraves & Ashley’s (2001) whole argumentation. We will focus on those aspects that are instructive with respect to our problem (P).

4. Case study 2: Real experimental reports in Holtgraves & Ashley (2001)

4.1. Setting up the starting p-context

In Holtgraves & Ashley (2001), the starting p-context includes Searle (1969), Austin (1962) and Searle & Vanderveken (1985) as direct sources. It also includes the result of Searle's thought experiment as summarised in (21)/(26) which the authors reformulate in the following way:

(29) “[...] utterances have both a propositional content and an illocutionary force (the speech act performed with the utterance).” (Holtgraves & Ashley 2001: 83)

The starting p-context is problematic in the sense of Subsection 2.5 in Kertész & Rákosi (this volume b):

- (30) a. “[...] it is not clear whether *comprehension* even involves *recognition* of the action that is performed with an utterance. That is, when we comprehend a speaker's remark, does that comprehension involve recognition, at some level, of the action that the speaker is undertaking? Does comprehension of ‘Don't forget to stop at the store’ involve recognition that the speaker is performing the act of reminding?
- b. Surprisingly, with the exception of research on recognition of indirect speech acts (e.g., Gibbs 1983), *very little empirical research* has addressed this issue.
- c. The purpose of this research was to undertake some initial analyses of *the extent to which comprehension involves recognition of the action* – or speech act — that is performed with an utterance.” (Holtgraves & Ashley 2001: 83; emphasis added)

As (30)(a) indicates, the p-context is informationally underdetermined. Searle's thought experiment (and speech act theory in general) simply presupposes in (4)(b) that speaker-intentions are recognised by the hearer without providing evidence for this assumption. Holtgraves and Ashley are not sufficiently content with the low plausibility value of this hypothesis to allow it to exist as a pure assumption. Moreover, Holtgraves & Ashley (2001: 84) refer to rival approaches, namely, “speech act theory for which illocutionary force recognition is not required” (see, e.g., Cohen & Levesque 1990). Thus, they raise the following problem:

(31) “[...] is it the case that addressees actually recognise the performance of a specific speech act?” (Holtgraves & Ashley 2001: 84)

Accordingly, the argumentation process should find out whether it is possible to assign a plausibility value to the following hypothesis:

(32) Comprehension involves recognition of the action that is performed with an utterance,

or to assign a plausibility value to its rival, namely, its negation, – i.e., whether (32) is plausible or implausible.

4.2. *The elaboration of the starting p-context*

The authors *extend the p-context* by new sources which are based on what they call “empirical investigations”. Although they emphasise that the problem in (31) has not yet been addressed by ‘empirical’ investigations, they refer to a series of similar problems which have been tackled with the help of psycholinguistic experiments. Thereby, they accept the following latent background assumption:

(33) $[0 < | \text{Real experiments are highly reliable data sources.} |_{HA} < 1]$

(33) exemplifies that the p-context may include methodological assumptions which significantly influence the plausibility of the experimental results. Since, however, the experiments referred to by Holtgraves & Ashley (2001: 84) bear only an indirect relationship to illocutionary force activation, they provide only weak evidence for (32). Their results are either coherent with (32) or they support (32) on the basis of some analogy between illocutionary force activation and, for example, drawing spontaneous trait inferences during the comprehension of a person’s behaviour. The analogy is established on the assumption that both are based on some kind of categorisation. Clearly, these pieces of evidence do not allow the authors to prefer (32) to its negation and make a decision between them.

Accordingly, the task they set themselves is “to investigate whether or not illocutionary force is activated during comprehension” (ibid.) by carrying out real experiments. Thus, the authors carry out four real experiments.

4.2.1. *Experiment 1*

Regarding the experimental design,¹⁴ participants were presented scenarios followed by a dialogue. For example:

(34) “Jenny and Emily had been close friends since grade school. Now they were rooming together at college.

Emily was very forgetful.

Today, Jenny was sure Emily *didn’t remember* (had forgotten) her dentist appointment.

Jenny: *Don’t forget* (I’ll bet you forgot) to go to your dentist appointment today.¹⁵

Probe: remind.” (Holtgraves & Ashley 2001: 85)

The task of the participants was to decide whether the probe word – in (34) the speech act verb *remind* – had been literally mentioned. Holtgraves and Ashley put forward the following conjectures about the possible outcome of the experimental procedure:

(35) a. “If comprehension of a remark involves activation of its illocutionary force, then, when the probe represents the speech act just performed, participants’ ability to verify that the probe had not been literally present should be slowed; activation of the speech act term should interfere with performance on this task. For example, participants should be slower at verifying that the word *remind* had not literally been present in the remark *Don’t forget to go to your dentist appointment today*

¹⁴ We do not reconstruct the experimental designs completely in either case. For further details of these, see the relevant passages in Holtgraves & Ashley (2001).

¹⁵ “The speech act version contained the italicized material; the control version was created by replacing the italicized material with the material in parentheses” (op. cit.).

relative to a control remark *I'll bet you forgot to go to your dentist appointment today.*

- b. On the basis of speech act theory and the empirical research cited above, we expected reaction times to be slowed, and errors to be increased, when the probe word characterized the speech act performed with the final remark than when it did not.” (Holtgraves & Ashley 2001: 85)

The participants of the experimental procedure were 55 students of psychology, all native speakers of English. They received a set of scenarios consisting of six speech act scenarios and six control scenarios. In the case of the speech act versions, the final sentence performed the particular speech act referred to by the probe word but this was not the case with the control scenarios. The correspondence between the intended speech act and the probe word was controlled with the help of a pretest. In the pretest, 75% of participants named the expected term correctly. The probe word never appeared in the text of the final remark of the scenarios. Nevertheless, in order to prevent such expectations, 12 filler scenarios were used which contained the probe word. Several details of the experimental situation such as the software used, the instructions, the presence of practice trials, the timing of the different stages of the completion of the required tasks etc. were controlled for, and described in the experimental report.

The perceptual data obtained had the structure “The time between the appearance of the probe word *W* on the screen and the keystroke by person No. *P* in item No. *I* was *x* ms.” Then, the perceptual data underwent authentication and interpretation. For example, perceptual data related to an incorrect decision about the presence/absence of the probe word were excluded from the analyses. Statistical analyses were conducted with both participants and stimuli as random variables. As a result, the following experimental data were obtained and made public:

- (36) a. $0 < | \text{Mean probe reaction time (and error rate) in speech act sentences was 1,034 ms (6.36\%)} \text{ in Experiment 1.} |_{Exp1} < 1$
 b. $0 < | \text{Mean probe reaction time (and error rate) in control sentences was 960 ms (1.52 \%)} \text{ in Experiment 1.} |_{Exp1} < 1$
 c. $0 < | \text{“Participants were significantly slower at verifying that the probe had not been literally present when it represented the speech act performed with the final remark than when it did not.”} |_{Exp1} < 1$

It is very important to realise that these data *are not true with certainty*. The experimental procedure, the interpretation and the authentication of the perceptual data may be faulty and be in need of modification. They cannot be totally controlled; it is always possible that relevant factors which influence the outcome of the experimental process to some extent remain unrevealed. From this it follows that experimental data can be only plausible.

As a next step, Holtgraves and Ashley *confronted* these data *with one of the rival hypotheses*, namely, with (32). With this aim in view, first they referred to the conjecture formulated in (35):

- (37) $0 < | \text{“If comprehension of a remark involves activation of its illocutionary force, then, when the probe represents the speech act just performed, participants’ ability to verify that the probe had not been literally present should be slowed.”} |_{HA} < 1 (= (35))$
 $0 < | \text{When the probe represented the speech act just performed, participants’ ability to verify that the probe had not been literally present was slowed.} |_{Exp1} < 1 (= (36))$

 $0 < | \text{Comprehension of a remark involves activation of its illocutionary force.} |_{I(37)} < 1$

Although this inference is not deductively valid, it is capable of making its conclusion plausible, since it can be shown that it belongs to the third group of plausible inferences (see Kertész & Rákosi, this volume b: Section 2.3):

(38) $0 < |\text{Comprehension of a remark involves activation of its illocutionary force.}|_{I(37)} < 1$

As (39) indicates, (38) supports (32):

(39) $0 < |\text{If comprehension of a remark involves activation of its illocutionary force, then comprehension involves recognition of the action that is performed with an utterance.}|_{HA} < 1$
 $0 < |\text{Comprehension of a remark involves activation of its illocutionary force.}|_{I(37)} < 1$

 $0 < |\text{Comprehension involves recognition of the action that is performed with an utterance.}|_{I(40)} < 1 (= (32))$

As for the *coordination and modification of the p-context*, experimental data originating from Experiment 1 as a data source provide support for (32). This means that the hypothesis that comprehension involves recognition of the action that is performed with an utterance has become plausible. The current state of the p-context is as in Figure 7.

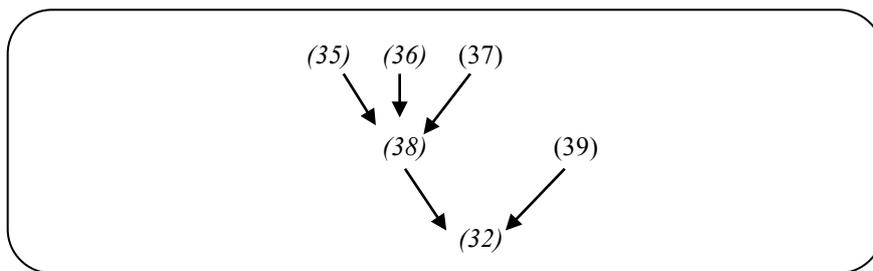


Figure 7: The state of the p-context after conducting Experiment 1

Nevertheless, the authentication of perceptual data raised some concerns about the reliability of the results. By analysing the data set, the authors revealed a possible systematic error which may have distorted the results:

(40) “A potential problem with this experiment is that participants took longer to read the final remark when it performed the relevant speech act [...] than when it did not [...]. Hence, slowed judgement speeds following the relevant speech act utterances could have been due to the greater comprehension difficulty for these remarks, relative to the control remarks.” (Holtgraves & Ashley 2001: 86)

In (37), the authors supposed that slowed reaction times indicate the presence of illocutionary force activation. After extending the p-context by (40), however, (37) had to be transformed into another inference which takes into consideration the presence of a second possible explanation of the experimental data as well. Thus, the following inference presents itself:

(41) $0 < |\text{If the probe represents the speech act just performed, and participants' ability to verify that the probe had not been literally present slows, while there is no difference in the comprehension difficulty for the final sentences of the speech act scenarios and the control scenarios, respectively, then comprehension of a remark involves activation of its illocutionary force.}|_{HA} < 1 (= (40))$

$0 < |$ When the probe represented the speech act just performed, participants' ability to verify that the probe had not been literally present was slowed. $|_{Exp1} < 1 (= (36))$
 $|$ [There was no difference in the comprehension difficulty for the final sentences of the speech act scenarios and the control scenarios, respectively. $| = ?]$

 $|$ Comprehension of a remark involves activation of its illocutionary force. $|_{I(41)} = ?$

The latent background assumption of (41), which was put between '[' and ']' above, has been *retrospectively and prismatically re-evaluated* in the sense of Subsection 2.6.1 of Kertész & Rákosi (this volume b). It had neutral plausibility during Experiment 1. Subsequent analysis of the experimental data, as (40) shows, however, questioned the assumption and showed that it was implausible. As a consequence, the premises of (41) are no longer capable of supporting the conclusion in the given state of the p-context, even though their plausibility did not change. Therefore, (38) loses its earlier plausibility value. This does not mean, of course, that the conclusion became implausible.

In order to find out whether the data originating from Experiment 1 can be used to support (32), a control experiment was proposed and a new subcycle initiated.

4.2.2. Subcycle 1: Experiment 2

Subcycle 1 of the argumentation process corresponds to Experiment 2. It replicates the first experiment with two *modifications* in the experimental design. First, the text of the final remarks of the control scenarios was lengthened. As a consequence, the latent background assumption of (41) was no longer implausible. Second, participants had a further task: from 24 words they had to choose those which described the remarks they read during the experiment. 6 words referred to speech acts which were performed with them in one of the scenarios, 6 words were irrelevant speech act terms, while 12 words were fillers.

There were two sets of perceptual data. The first set consisted of data with a similar structure as had been the case with Experiment 1: "The time between the appearance of the probe word *W* on the screen and the keystroke by person No. *P* in item No. *I* was *x* ms." The second set of perceptual data had the structure "Person No. *P* hit the button *Y/N* in item No. *I*." The following experimental data were obtained after the interpretation and authentication of the perceptual data:

- (42) a. $0 < |$ Mean probe reaction time (and error rate) in speech act sentences was 1,099 ms (9.42%) in Experiment 2. $|_{Exp2} < 1$
 b. $0 < |$ Mean probe reaction time (and error rate) in control sentences was 1,047 ms (5.44 %) in Experiment 2. $|_{Exp2} < 1$
 c. $0 < |$ "First, the participants made more errors when the probe represented a performed speech act than when it did not [...]. And second, the participants were again significantly slower at making this judgement when the probe represented a prior speech act than when it did not [...]." $|_{Exp2} < 1$

(42) provides support for (38) in a similar manner as (36) did:

- (43) $0 < |$ If the probe represents the speech act just performed, and participants' ability to verify that the probe had not been literally present slows, while there is no difference in the comprehension difficulty for the final sentences of the speech act scenarios and the control scenarios, respectively, further participants make more errors when the probe represented a performed speech act than when it did not, then comprehension of a remark involves activation of its illocutionary force. $|_{HA} < 1 (= (40))$

$0 < |$ When the probe represented the speech act just performed, participants' ability to verify that the probe had not been literally present was slowed. $|_{Exp2} < 1 (= (42))$
 $0 < |$ The participants made more errors when the probe represented a performed speech act than when it did not. $|_{Exp2} < 1 (= (42))$
 $0 < |$ There was no difference in the comprehension difficulty for the final sentences of the speech act scenarios and the control scenarios, respectively. $|_{Exp2} < 1$

 $0 < |$ Comprehension involves recognition of the action that is performed with an utterance. $|_{H(40)} < 1 (= (32))$

Moreover, (42) indicates that the concern in (40) is unfounded. Thus, (41) can be rejected and (37) can be added to the p-context again. Figure 8 shows the current shape of the p-context.

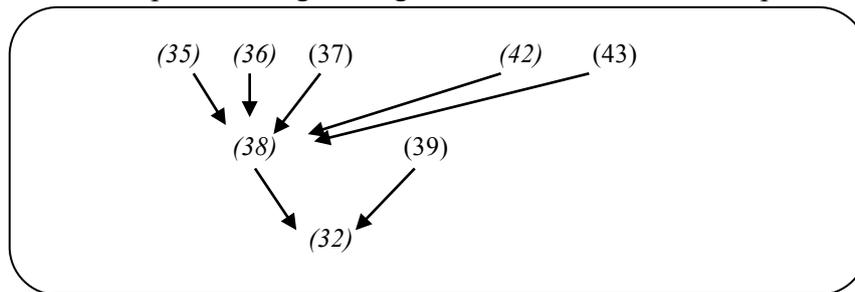


Figure 8: The state of the p-context after conducting Experiment 2

The authentication of the perceptual data, however, indicated another possible error source. Namely, the authors found that the differences in reaction times were significant only over participants but not over items in both experiments. Therefore, a retrospective re-evaluation of the results was needed and a second subcycle was initiated in order to find out whether there is a relevant, but unrealised factor which biases the results and makes the proposed interpretation faulty.

4.2.3. Subcycle 2: Experiment 3

The purpose of Experiment 3 was to check whether the results of the first two experiments were influenced by factors different from those which have been assumed. More precisely, it might have been the case that the results which supported (32)

[...] were due to some type of context checking procedure that occurred when the participants verified the probes. That is, reaction times might have been slowed for the speech act probes, not because these words had been activated during comprehension, but because when the judgments were made, the speech act probes were more plausible than the control words. (Holtgraves & Ashley 2001: 87)

To filter out such factors, Experiment 3 was carried out. Although the stimulus material remained unchanged, the participants had to solve another task. Instead of indicating whether or not the probe word literally occurred in the conversation they were presented, they had to check whether a certain letter string was a word. Holtgraves and Ashley put forward the following conjecture about the possible outcome of the experimental procedure:

(44) $0 < |$ “If comprehension of the utterance involved speech act activation, the participants should be significantly faster at this task when the target represented the performed speech act than when it did not.” $|_{HA} < 1$

The interpretation and authentication of the perceptual data led to the following set of experimental data:

- (45) (a) $0 < | \text{Mean lexical decision reaction time in speech act sentences was 1,002 ms in Experiment 3.} |_{Exp3} < 1$
- (b) $0 < | \text{Mean lexical decision reaction time in control sentences was 1,095 ms in Experiment 3.} |_{Exp3} < 1$
- (c) $0 < | \text{In error-free trials, the overall error rate was 9.6\%, and this rate did not vary across conditions in Experiment 3.} |_{Exp3} < 1$
- (d) $0 < | \text{“[...] the participants were significantly faster at the lexicon decision task when the target represented the prior speech act than when it did not [...]” in Experiment 3.} |_{Exp3} < 1$

It this case, the results were significant both over items and participants. Thus, the following plausible inference indicates the *confrontation* of these experimental data with the previous conjectures:

- (46) $0 < | \text{“If comprehension of the utterance involved speech act activation, the participants should be significantly faster at this task when the target represented the performed speech act than when it did not.”} |_{HA} < 1 (= (44))$
 $0 < | \text{“[...] the participants were significantly faster at this task when the target represented the performed speech act than when it did not.”} |_{Exp3} < 1 (= (45))$

 $0 < | \text{Comprehension of the utterance involved speech act activation.} |_{I(46)} < 1$

The conclusion of (46) provides support for (32) again. Changes in the p-context after extending it with the results of Experiment 3 are as in Figure 9.

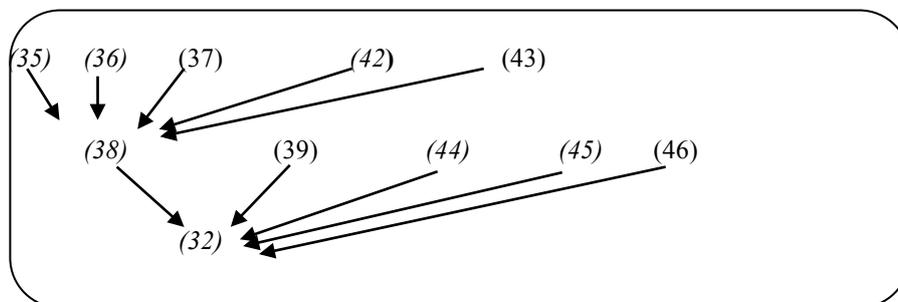


Figure 9: The state of the p-context after conducting Experiment 3

Nevertheless, the reliability of Experiment 3 as a data source can be questioned according to Holtgraves & Ashley (2001: 87), since it is possible that in Experiment 3 the reason for faster lexical decisions was not the participants’ recognising the speech act at issue, but the priming of semantic associates of words. Thus, we are in a similar situation as in the case of Experiment 1: (46) has to be retrospectively re-evaluated. In the light of new information, (46) is transformed into an inference which is not capable of providing support for the hypothesis about speech act activation. Namely, since no plausibility value can be assigned to the latent background assumption of (47), it is not capable of providing support to its conclusion:

- (47) $0 < |$ If the participants are significantly faster when the target represents the performed speech act than when it does not, and there is no semantic priming effect, then comprehension of the utterance involves speech act activation. $|_{HA} < 1$
 $|$ [There was no semantic priming effect. $| = ?]$
 $0 < |$ "[...] the participants were significantly faster at this task when the target represented the performed speech act than when it did not". $|_{Exp3} < 1 (= (45))$

 $|$ Comprehension of the utterance involved speech act activation. $|_{I(47)} = ?$

Therefore, in order to eliminate possible priming effects, the experimental design was modified again.

4.2.4. Subcycle 4: Experiment 4

In a new subcycle, Experiment 4 replicated Experiment 3, the only difference being that the sentences performed were equated with their semantic associates. That is, the text of the control sentences and the corresponding speech act sentences has been made as similar as possible. The authors obtained the following experimental data:

- (48) a. $0 < |$ Mean lexical decision reaction times in speech act sentences was 887 ms in Experiment 4. $|_{Exp4} < 1$
 b. $0 < |$ Mean lexical decision reaction times in control sentences was 999 ms in Experiment 4. $|_{Exp4} < 1$
 c. $0 < |$ In error-free trials, the overall error rate was 9%, and this rate did not vary over conditions in Experiment 4. $|_{Exp4} < 1$
 d. $0 < |$ "[...] the participants were significantly faster at the lexicon decision task when the target represented the prior speech act than when it did not [...]" in Experiment 4. $|_{Exp4} < 1$ (Holtgraves & Ashley 2001: 87f.)

Since the results of Experiment 4 are in harmony with those of Experiment 3, these data not only provide further support for (32) but they also motivate adding (46) to the p-context again. Therefore, the outcome of the former cannot be due to any priming effect. This means that (47) can be retrospectively re-evaluated and transformed into the following plausible inference:

- (49) $0 < |$ If the participants were significantly faster at the lexicon decision task when the target represented the prior speech act than when it did not, and there is no semantic priming effect, then comprehension of the utterance involves speech act activation. $|_{HA} < 1$
 $0 < |$ There was no semantic priming effect. $|_{Exp4} < 1$
 $0 < |$ "[...] the participants were significantly faster at the lexicon decision task when the target represented the prior speech act than when it did not [...]" in Experiment 4. $|_{Exp4} < 1 (= (48))$

 $0 < |$ Comprehension of the utterance involved speech act activation. $|_{I(49)} < 1$

After adding the data obtained in Experiment 4, the p-context is as in Figure 10.

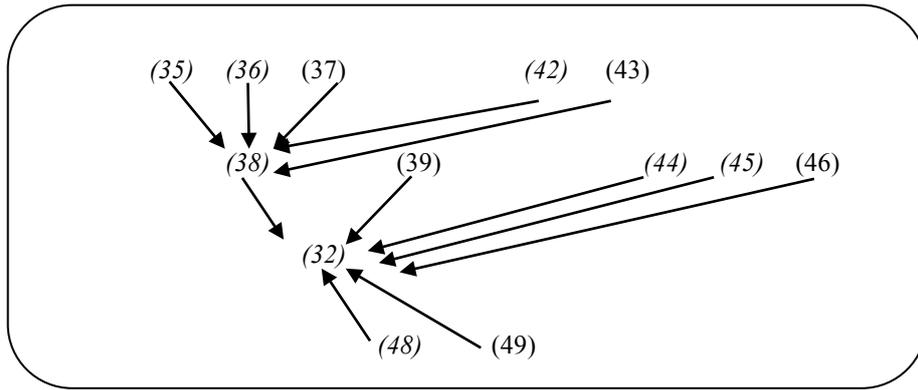


Figure 10: The state of the p-context after conducting Experiment 4

4.3. The comparison with other solutions

The p-context includes the methodological norm which attributes high reliability to real experiments as the source of the plausibility of hypotheses. Therefore, Experiments 1-4 as sources make it possible to assign a very high plausibility value to (32): “[...] we have converging evidence from four separate experiments and three different methods that illocutionary force is activated” (Holtgraves & Ashley 2001: 87f.).

The authors explicitly compare their result – i.e. (32) along with its increased plausibility value gained from the four real experiments – to *two rival hypotheses* and, accordingly, apply the Contrastive Strategy. The first is Cohen and Levesque’s (1990) claim according to which

(50) “[s]peech act recognition is not required, although it may be entailed in the recognition process.” (Holtgraves & Ashley 2001: 87f.)

The second hypothesis is the stance adopted by conversation analysis as put forward for example in Schegloff (1991) which says that

(51) “[...] an interactant’s understanding of what has been accomplished with a remark is revealed in that person’s subsequent turn at talk; there is no need to consider a dimension of illocutionary force.” (ibid.)

Experimental data from Experiments 1-4 are not only weak evidence for (32) but also against its rivals. Thus, (36), (42), (44) and (48) provide *strong evidence* for (32) and against its rivals, (50) and (51), as well (see Kertész & Rákosi, this volume b: Section 3.2).

4.4. The final p-context

The final p-context includes (32) as the *resolution* of the problem raised in (31) on the basis of the information available. It is the resolution of (31) in the sense of Subsection 2.5 of Kertész & Rákosi (this volume b), because among the solutions (50), (51) and (32), it is (32) that has been assigned the highest plausibility value and its acceptance leads to the least p-problematic p-context – even if this is merely provisional and can be retrospectively re-evaluated in later cycles of the argumentation process.

However, with respect to the relationship between thought experiments and real experiments in pragmatics, Holtgraves and Ashley's experiments work as a prism through which the outcome of Searle's thought experiment can be re-evaluated retrospectively:

- (52) $0 < | \text{Comprehension involves recognition of the action that is performed with an utterance.} |_{HA \& \text{ Experiments 1-4}} < 1$ (= (32))
 $0 < | \text{If comprehension involves recognition of the action that is performed with an utterance, then "[...] utterances have both a propositional content and an illocutionary force (the speech act performed with the utterance)."} |_{HA} < 1$
 $[0 < | \text{If "[...] utterances have both a propositional content and an illocutionary force (the speech act performed with the utterance)", then the analysis of illocutionary acts must capture both the intentional and conventional aspects of meaning and especially the relationship between them.} |_{HA} < 1]$

 $0 < | \text{The analysis of illocutionary acts must capture both the intentional and conventional aspects of meaning and especially the relationship between them.} |_{I(52)} < 1$
 (= (28))

This also means that experimental data originating from the four experiments conducted by Holtgraves and Ashley considerably increased the plausibility value of (21) as well.

5. The relationship between thought experiments and real experiments

5.1. Introductory remarks

On the basis of the analyses presented in Sections 3 and 4, we have reached the point where different aspects of the multifaceted relationship between Searle's thought experiment and the experiments presented by Holtgraves and Ashley can be revealed. These findings can be *generalised* in order to obtain an answer to problem (P). We will do this in three steps:

Step 1: The analyses of the thought/real experimental reports suggest the generalisation that thought experiments and real experiments may be continuous in several respects.

Step 2: On the basis of our analyses in Sections 3 and 4, we can realise that there is a striking similarity between the argumentation structures of Searle's thought experimental report and Holtgraves & Ashley's experimental report. It seems to be instructive to examine possible analogies between the argumentation processes presented in thought experimental reports and in reports on real experiments.

Step 3: Analogies between the argumentation structure of thought experimental reports and real experimental reports might indicate a deeper similarity between real and thought experiments themselves. That is, it is possible that there is a basic similarity between the functioning of both experimental processes, too. Of course, the question is whether and to what extent the experimental processes can be reconstructed from the experimental reports.

5.2. Step 1: The continuity between thought experiments and real experiments

On the basis of the analyses we presented in Sections 3 and 4, there seem to be *three kinds of continuity* between thought experiments and real experiments in the cases we examined.

First, the starting p-context of Experiment 1 also includes the result of Searle's thought experiment as summarised in (28):

(28) The analysis of illocutionary acts must capture both the intentional and conventional aspects of meaning and especially the relationship between them.

Moreover, as we have noticed in Section 4.1, the starting p-context of Holtgraves & Ashley (2001) includes hypotheses whose direct sources are, alongside Searle's thought experiment, further seminal contributions to speech act theory. We have also seen that the experimental designs of Experiments 1-4 were built upon the application of speech act theory. Namely, in the text of the scenarios, probe words which characterised the given speech act were chosen on the basis of the application of speech act theory.¹⁶ Since the classical contributions to speech act theory made use of thought experiments (cf. Section 3), thought experiments are – although not solely – sources of the plausibility of certain premises used in real experiments. Accordingly, we are entitled to put forward the generalisation that the *results of thought experiments also contributed to the plausibility of the result of Experiments 1-4* in an indirect way.

Second, Searle's thought experiment has taken the existence of the illocutionary aspects of meaning for granted and focused solely on the presence of the conventional aspects.¹⁷ Conversely, Holtgraves and Ashley tried to find experimental evidence for the presence of the illocutionary force of utterances but did not investigate their conventional aspect. Since all four experiments they conducted provided evidence for the existence and recognition of speaker intentions, they considerably increased the plausibility of that part of (28) which is related to the intentional aspects of utterances. In this way, the *results of Experiments 1-4 also contributed to the plausibility of the result of Searle's thought experiment*.

To sum up, Holtgraves and Ashley's experiments work as a prism through which the outcome of Searle's thought experiment can be re-evaluated retrospectively. At the same time, however, Experiments 1-4 relied on hypotheses which were partially and indirectly supported by thought experiments. It is important to emphasise that this process, as a result of which the finding of Holtgraves and Ashley's real experiments feeds back to that of Searle's thought experiment and increases its plausibility, is not circular. Thus, the generalisation is that the result of real experiments may increase the plausibility of the result of thought experiments in the course of the cyclic, retrospective and prismatic re-evaluation of information: new pieces of information were made use of, and the thought experiment and the real experiment were conducive to the creation of a more comprehensive picture by unifying their virtues.

There is also a third kind of relationship between Holtgraves and Ashley's experiments and thought experiments. Namely, *the experimental designs of Experiments 1-4 (and in general, every experimental design) can be interpreted as a special kind of thought experiment*. This interpretation is based on the circumstance that experimenters go through the experimental process in their mind in order to test whether the procedure proposed will provide relevant and reliable data. Of course, the results of such thought experiments are

¹⁶ The decisions were checked in a pretest. Despite this, the primary source of the plausibility of several hypotheses on which the experimental design was based, is speech act theory – the pretest only increased these values.

¹⁷ See also Section 5.2.

considerably vaguer and less reliable than those of the real experimental procedure. Despite this, the experimental design is an indispensable part of real experiments. It delineates and motivates the details of the experimental procedure to be carried out, and it also provides several points of reference against which the results and all facets of the experimental process can be interpreted and checked in order to gain reliable data.

5.3. Step 2: Analogies between the thought experimental reports and experimental reports

In Sections 3 and 4, we reconstructed one of Searle's thought experimental reports and Holtgraves & Ashley's experimental report about four real experiments with the help of the p-model. In both cases, we have found that the presentation of the thought experiment and the real experiments can be interpreted as *plausible argumentation processes*. These argumentation processes show a *strong analogy* at several points:

- The starting p-contexts included some theory of *meaning*. Nevertheless, the plausibility of some basic hypotheses was *questioned*.
- In order to find out whether the questioned hypotheses are tenable, they were *tested under controlled circumstances*: in the case of Searle's thought experiment, an imagined situation was described and analysed, with Holtgraves and Ashley's real experiments, a group of native speakers was put into an experimental setting in order to monitor their linguistic behaviour.
- Controlled circumstances were, in contrast to naturally emerging situations, *carefully designed and created* with the aim of gaining relevant, reliable and unbiased information.
- *Conjectures* were put forward summarising the expectations of the researcher(s) about possible consequences of the created/imagined state of affairs.
- A *chain of plausible inferences* was made in both cases. These inferences connected the directly emerging pieces of information from the given situation with higher level hypotheses which summarised the results and could be confronted with the hypotheses of the theory (i.e., compared to the previous conjectures).
- The hypotheses proposed were *reformulated* during the (thought) experimental process and were gradually replaced by refined, re-evaluated versions.
- Nevertheless, the outcome of the experimental procedure and the picture emerging from the imagination of the described situation, respectively, were not totally reliable but only *plausible* to a certain extent. In the case of real experiments, control experiments were conducted to rule out possible systematic errors, while Searle's thought experiment made use of an analogous situation as well as different possible interpretations of the same situation.
- In both cases, the argumentation process was *cyclic*, not linear, in order to make some sources more reliable.
- Therefore, one of the most substantial characteristics of the argumentation processes was that they operated with *uncertain, fallible and often conflicting pieces of information* and *have led not to certainly true, but only plausible results*.
- Several details of the experimentation process/of the imagination of the described situation were *missing* from the experimental report. A concise and edited version replaced the original, real event and provided an *informationally reduced and/or modified picture* of the experimental procedure.
- The *evaluation* of experimental results, however, could only start from the experimental report which contained elements or traces of the original argumentation process, as well as argumentative tools needed to make this reduced set of information coherent, comprehensible and persuasive for the reader.

Of course, there are also *differences* between the argumentation structures of Searle's thought experimental report and Holtgraves and Ashley's real experimental report, since analogy means solely (structural) similarity but not identity. The most important difference is perhaps that Holtgraves and Ashley's argumentation is of a more cyclic character than Searle's, since the four experiments presented are built on each other and provide revised versions of the previous one, while in Searle's thought experimental report, the central hypothesis is elaborated gradually, that is, it is re-evaluated during the argumentation process.

Although there is a strong analogy between the argumentation structures of Searle's thought experimental report as well as Holtgraves and Ashley's experimental report, two concerns emerge as a result of the above survey. First, from the analogy between the argumentation structures, one must not automatically conclude that there is also an analogy between the thought experimental process and the real experimental process itself. It may be the case, for example, that thought experiments only imitate the style of real experiments by applying the rhetorical tools of their experimental reports, while there are considerable and deep differences between the two kinds of experimental processes themselves. Second, it is questionable whether, and to what extent, these experimental processes can be reconstructed from the reports about them. A sufficiently complete reconstruction can only be achieved if reports contain all information which is relevant for the identification and evaluation of the steps of the experimental process. This means that the reader should be in a position to reconstruct and continue the non-public argumentation process. In this way, the reader could become a *virtual participant in the creation, analysis and evaluation of the thought/real experimental data*. Therefore, in the next subsection we will examine whether this is the case and try to reconstruct and compare the thought experimental and the real experimental process.

5.4. Step 3: Analogies between real and thought experiments

5.4.1. The structure of real experiments¹⁸

The *experimental design* of a real experiment is a comprehensive preliminary description of all facets of the process of experimentation allowing for a delineation of the events which are supposed to occur and a rough estimation of the results. It is therefore nothing else but a *thought experiment* which is, of course, less informative than the real experiment. The outcome of the experimental procedure always contains new pieces of information which may strengthen (and make more precise) but also weaken or refute this preliminary guess.

The experimental designs of the four experiments were detailed and basically well-planned. Nevertheless, there were a few problems. The most important concern is typical of psycholinguistic experiments: there is a considerable amount of information loss because the experimental report presents the text of only one scenario. Therefore, the appropriateness of the stimulus material used in the experiments cannot be checked by the reader of the experimental report. Second, the change of the text of the control scenarios in Experiment 4 in comparison to Experiment 3 is not clear (cf. Holtgraves & Ashley 2001: 87). Third, measuring the reaction times during the task of deciding whether a letter string is a word or a non-word counts as a method for avoiding lexical priming effects (cf. Keysar et al. 2000: 589). It was, however, precisely lexical priming effects which Holtgraves and Ashley referred to as possible sources of error in Experiment 3 which made use of this method.

¹⁸ We have already proposed a model of psycholinguistic experiments based on the p-model (cf. Kertész & Rákosi 2012, Rákosi 2011a,b, 2012), by relying on insights in the current literature on scientific experiments. Therefore, we will apply this model to the reconstruction of the structure of Experiments 1-4.

The *experimental procedure* is a material procedure where an experimental apparatus is set up and its operation is monitored and recorded under controlled circumstances. During the experimental procedure *perceptual data*¹⁹ are gained which undergo *authentication* and *interpretation*. Authentication means that the experimenter has to check whether sources of noise do not distort the results, that is, whether they have been ruled out successfully or, if this is not possible, whether their effect can be eliminated with the help of statistical methods. As a result of this process, *experimental data*²⁰ are obtained which are then confronted with the given theory or with some rival theories. Nevertheless, it may happen that the interpretation and authentication of the perceptual data indicates shortcomings in the experimental procedure, in the experimental design, or in the theoretical model of the phenomena or of the apparatus. If things do not run smoothly, one turns back to some earlier stage of the experimentation process and modifies a component until there is mutual support among the constituents. All facets of the experiment have to be *re-examined*.

The experimental report does not contain the set of perceptual data, as is the case with most psycholinguistic experiments on cognitive metaphors. Thus, their interpretation and authentication cannot be properly judged by the reader. This circumstance – together with the absence of the stimulus material – decreases the plausibility of the experimental data. On the other hand, Holtgraves and Ashley revealed and checked several possible systematic errors with the help of control experiments. This makes their results more plausible, because in this way the reliability of their experiments increased. Nevertheless, the authentication can never be complete. In this case, for example, the following problematic points can be mentioned:

- In contrast to most psycholinguistic experiments, where feedback is used only in practice trials, participants received feedback about the correctness of their decision and their reaction times. This might have influenced the participants' decisions.
- The description of the statistical tools applied is rather sparse.
- Experiment 2 was intended to be a control experiment for Experiment 1. Nevertheless, the experimental procedure was not only changed at the point which was deemed to be relevant for preventing a possible systematic error (lengthening the final remarks). It was also indicated that the participants will receive a second task as well. This might have influenced the focus of attention and the behaviour of participants.
- The authors did not pay attention to the great, and probably significant difference between error rates in Experiment 1 and Experiment 2.
- Error rates were high (above 9%) in all experiments.
- There was a significant difference between decision times in Experiments 3 and 4 both in speech act and in control scenarios. This is surprising because the tasks in the speech act scenarios were completely identical in the two cases.

From these considerations it is clear that the experimental data gained are not true with certainty but *only plausible* on the basis of these experiments as sources. Their confrontation with the hypotheses of the theory, with previous conjectures, involves factors which indicate that experimental results are not decisive. They *cannot prove or falsify* hypotheses but are capable of increasing their plausibility or making them implausible as weak or strong evidence for/against them. For example, the reconstruction of the argumentation structure of

¹⁹ In accordance with our definition of the notion of 'datum' (Kertész & Rákosi, this volume b: Section 3.1), by 'perceptual data' we mean plausible statements the plausibility value of which stems from perceptual information as a direct source.

²⁰ Likewise, by 'experimental data' we mean plausible statements the plausibility value of which stems from an experiment as a direct source.

Holtgraves & Ashley's experimental report contained the following plausible inference (cf. Section 4.2):

- (37) $0 < |$ “If comprehension of a remark involves activation of its illocutionary force, then, when the probe represents the speech act just performed, participants' ability to verify that the probe had not been literally present should be slowed.” $|_{HA} < 1$ (= (35))
 $0 < |$ When the probe represented the speech act just performed, participants' ability to verify that the probe had not been literally present was slowed. $|_{Exp1} < 1$ (= (36))

 $0 < |$ Comprehension of a remark involves activation of its illocutionary force. $|_{I(37)} < 1$

After the (partial) reconstruction of the experimentation process, it might become more evident that the first premise of this inference is not true with certainty; it is a long way from the antecedent to the consequent. Similarly, our reconstruction has shown that there are many factors which might influence the plausibility of the latent background assumptions which would be needed to transform this inference into a deductively valid one (cf. Kertész & Rákosi, this volume b: Section 2.3). From this it follows that the experiments could only provide partial support for the acceptance or rejection of the hypotheses of the given theory.

To sum up our analyses, the p-model interprets experiments as *cyclic processes*. This cyclic process is organised and conducted by a *plausible argumentation process*. This process governs the relationship among hypotheses of the experimental design, the theoretical model of phenomena, the theoretical model of the experimental apparatus, the theory being tested and its rivals as well as statements describing the events of the experimental procedure, which capture the results of the interpretation and authentication of perceptual data etc. See Figure 11.²¹

²¹ In Figure 11 and 12, simple arrows indicate successive stages of the experimental process; dotted arrows signify the non-public argumentation process which organises the experimental process.

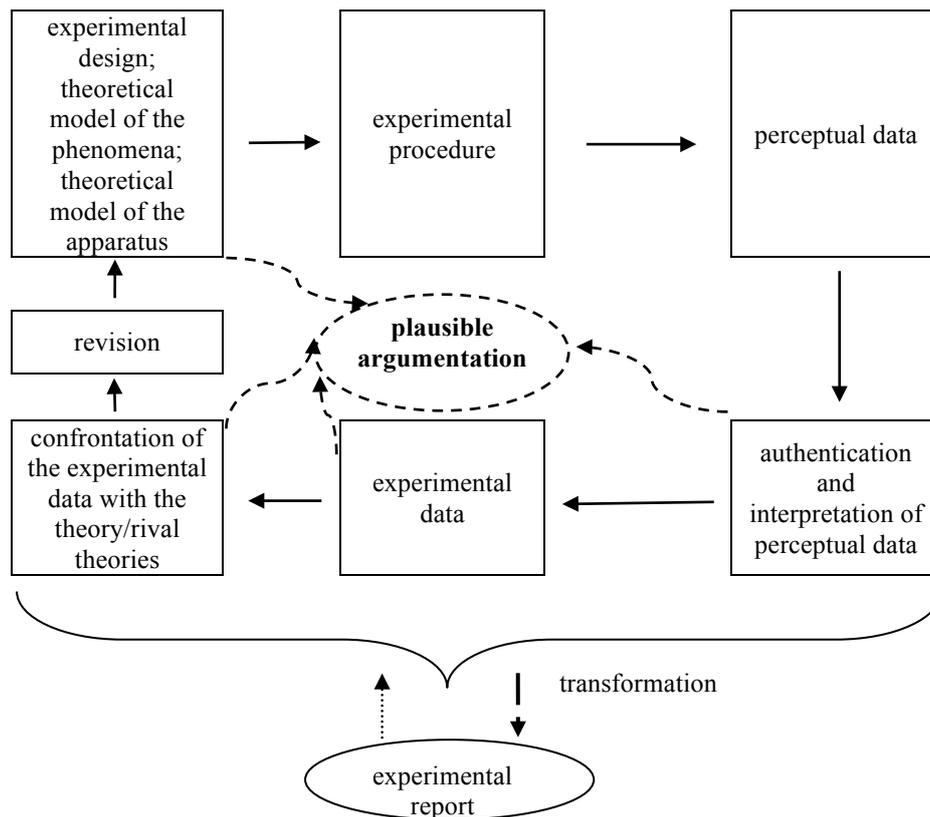


Figure 11: The structure of real experiments

The *effectiveness and comprehensiveness of the plausible argumentation process* organising the conduct and control of real experiments largely determines the reliability of the obtained results.

This argumentation process is, however, not public but remains a *private affair* of the experimenters. It is transformed into an experimental report/paper summarising the results of the experiment and making them available to the scientific community. Clearly, this transformation can be regarded as acceptable if it does not change the plausibility value of the hypotheses of the original argumentation. This is of utmost importance because the researcher runs the risk of eliminating relevant information from the published report so that important decisions remain outside public control, and she/he overestimates the plausibility of the results. From this it follows that the acceptability of the experimental report is also influenced by the reliability of the *transformation* of the non-public argumentation process into its public version.

5.4.2. The structure of thought experiments

Our next task is to examine whether this model of real experiments can be applied to thought experiments smoothly, in such a way that besides certain differences, we find a deep analogy between thought experiments and real experiments. We cannot, of course, perform this task comprehensively and in general. Therefore, we will try to find analogies between the stages of the model of real experiments presented in the previous subsection and the stages of Searle's thought experiment. If this is possible, then we can generalise our results step by step and the outcome of our considerations will be a model of thought experiments developed by analogy to our model of real experiments.

As we have seen in the previous subsection, with real experiments the starting point is the elaboration of the experimental design. This includes a comprehensive description of the *equipment* to be used and the *circumstances* under which the experimental procedure will take place, as well as a rough depiction of the *events* which are supposed to occur in this situation. The first step of Searle's thought experiment seems to be similar but somewhat simpler: it is the *delineation of a state of affairs (situation)* which has to be detailed enough to be easily *imagined* (visualised or simulated) by the members of the scientific community. Our reconstruction of Searle's thought experimental report in Section 3.1 interpreted the passage cited in (4)(a) as the description of a situation in the above sense. Thus, an important common feature is that both real and thought experiments intend to *create a situation* from which relevant and reliable information can be gained.

The second component of real experiments is a *material procedure*, i.e. the operation and monitoring of the functioning of an experimental apparatus. With thought experiments, the *imagination (visualisation, simulation)* of the situation at issue seems to correspond to this part of the experimental procedure. It is a *cognitive procedure* which has to be performed both by the author and the reader. Of course, there is an important difference: while the material procedure takes place in the physical world, *partly* independently of the researcher, the cognitive procedure is performed in the head of the researcher. Nevertheless, the material procedure is *not totally* independent of the researcher, either, because it has been set up and controlled by him/her; and the theoretical commitments of the researcher may influence the way the situation is imagined. Moreover, a thought experiment cannot work without the active participation of the reader: he/she has in fact to imagine the situation described. Therefore, a simple reading and acceptance of Searle's narrative is not enough. The reader should "process" the situation by relying on his/her linguistic intuition and everyday knowledge. Similarly, the reader of an experimental report has to have some understanding of the functioning of the equipment used – although a thorough knowledge of it is usually not required.²² Therefore, the reader's control over the thought experimental procedure usually seems to be stronger than is the case with real experiments.

The third step of real experiments is the collection of the *perceptual data* obtained during the experimental procedure. In most cases this means reading the dial on the equipment and formulating the corresponding plausible statement. In Searle's thought experiment, this constituent is missing: it does not contain data whose plausibility would originate directly from contact with the physical world. Nevertheless, our everyday experience and linguistic intuition may allow us to make highly plausible statements about the events which should occur in the given situation on the basis of some analogy to similar real-world situations, that is, on our experiences. These will be termed the *experiential data* emerging from the situation investigated. For example, the data presented in (5) or in (10) in Section 3.2 are experiential data in this sense because they provide a description of an imaginary situation based on everyday linguistic intuition. That is, thought experiments rely on data which result from an *indirect contact* with the empirical world.

The next move in real experiments is the *authentication and interpretation of the perceptual data*. The experimenter checks whether the perceptual data are reliable enough, and describes them with the help of the concepts of some theory. Since the authentication involves the use of statistical tools, the obtained *experimental data* lose a direct connection to the physical world. As a consequence, they are not true with certainty. One cannot be sure that all possible sources of random and systematic errors have been ruled out: every check of the correct functioning of the experimental apparatus requires new measurements, which requires new authentication, and so on – the *experimenter's regress* (cf. Collins 1985) can be broken only in a more or less arbitrary manner. In thought experiments, the plausibility value

²² Thus, for example, theoretical physicists' and experimental physicists' knowledge of the working of an experimental apparatus may differ in depth.

of the experiential data depends, among other things, on the degree of the *typicality*, *transparency* and *clear-cut character* of the components of this state of affairs.²³ The more determined and unanimous the everyday experiences, the cultural, linguistic or other kinds of intuitions related to this state of affairs are, the higher the plausibility value which can be assigned to the experiential data. It may be the case, of course, that the reader does not share the thought experimenter's intentions and judges the experiential data less plausible or even implausible; it is also possible that the reader has other intuitions in connection with the depicted situation. Then, on the basis of the experiential data, the theoretical consequences of the events which are supposed to occur in the situation have to be identified. That is, the experiential data have to be described with the help of the conceptual apparatus of the theory. As a result, *thought experimental data*²⁴ are obtained. With Searle's thought experiment, hypothesis (7) is, for instance, a consequence of the situation analysed in (4)(a)-(c) which has been obtained on the basis of the experiential datum (5)(a). Of course, our reconstruction of Searle's argumentation process could have interpreted (7) as a thought experimental datum as well; i.e. if we analyse the relationship between (5)(a) and (7) not as a connection between a premise and a conclusion of the plausible inference in (6), but deem (7) to be the re-evaluation of (5)(a) with the help of the terminology of Gricean theory of meaning, then (7) will originate from a direct source and could be regarded as a datum. Similarly, (12) could also be interpreted as a thought experimental datum which is the re-evaluation of the experiential datum (10).

Following this, in both real and thought experiments, experimental data are confronted with the hypotheses of a theory or rival theories. If the results seem to be reliable, then the real/thought experimental process is transformed into a real/thought experimental report. Akin to real experimental reports, thought experimental reports are informationally impoverished in comparison to the thought experimental process conducted by the experimenter:

By the time the thought experiment is communicated it is in a polished form. That is, we are not shown the tinkering that went into setting it up and refining it. [...] When conveyed, a thought experiment is as packaged and polished as a real-world experiment is when it is published. (Nersessian 1993: 296)

The above characterisation of thought experiments supports the hypothesis that there is a strong parallelism between the stages of real and thought experiments. This correlation can also be found between the argumentative aspects of real and thought experiments. A further component of the analogy between real and thought experiments is that thought experiments are akin to real experiments insofar as the experimental process is guided by a *plausible argumentation process*. This argumentation process establishes the connections among hypotheses related to the components of the real/thought experiment and makes it possible to decide in which cases some component has to be revised because it is in conflict with other components or under what circumstances no revision is needed because there is a fit among the different elements. Moreover, with both real and thought experiments, this argumentation process is not made public. One can reconstruct it to some extent from the experimental report/paper which contains several traces of it. This means that the real/thought experimental process can be evaluated only through the prism of its informationally reduced and edited version. Reproducibility is in both cases a very important means of judging the reliability of the outcome of the experimental process: repetitions may reveal suppressed but relevant factors which indicate that the results are less reliable or even unreliable.

²³ Of course, the whole of the situation need not be typical – situations depicted in thought experiments are often fantastic or unrealistic, but contain realistic and well-known elements.

²⁴ In analogy to the notion 'experimental data', 'thought experimental data' should be understood as plausible statements the plausibility value of which stems from a thought experiment as a direct source.

In summary, the p-model interprets thought experiments by analogy with real experiments, that is, as *cyclic processes* organised and conducted by a *plausible argumentation process*.²⁵ Both real and thought experiments are characterised by a *double argumentative structure*: they are organised and guided by a non-public argumentation process and they are transformed into a public argumentation process (thought/real experimental report). See Figure 12.

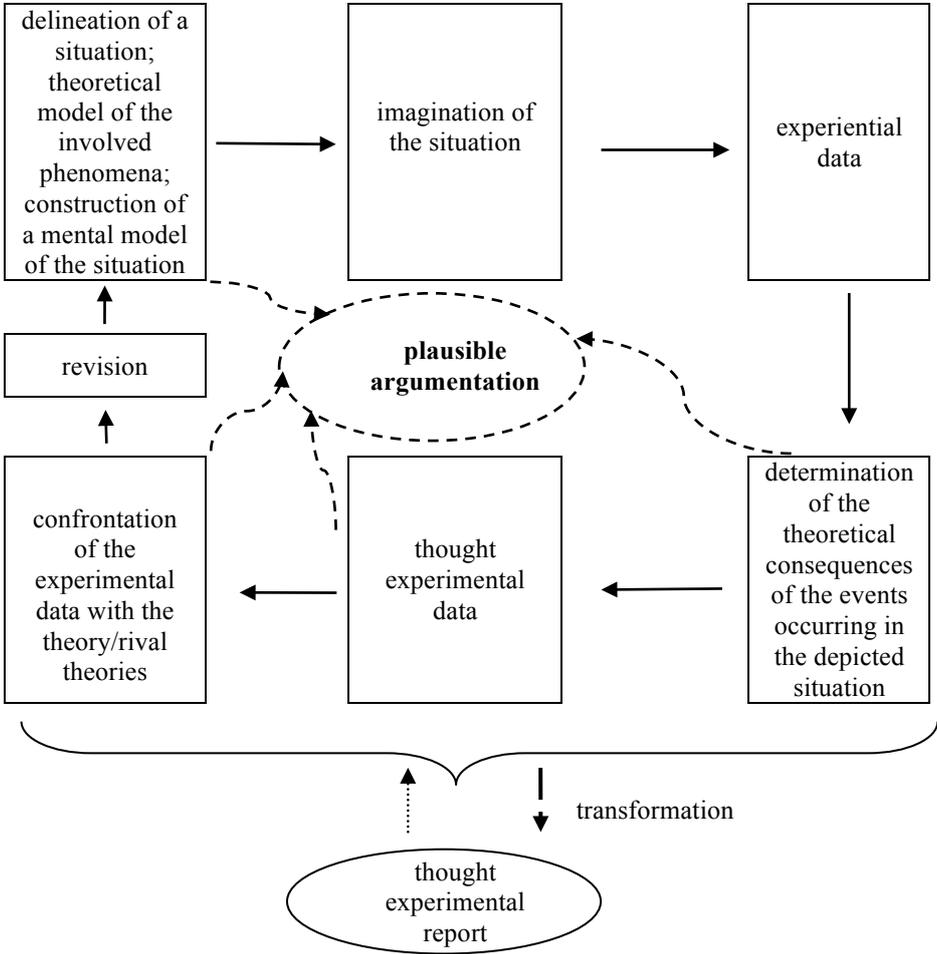


Figure 12: The structure of thought experiments

6. The solution to (P)

We can now summarise our solution to the problem (P) as follows:

- (SP) (a) The result of a thought experiment may be included in the starting p-context of a real experiment and accordingly, there may be an indirect continuity between them. Such thought experiments may serve as sources of the plausibility of the

²⁵ It is worth remarking that in Holtgraves and Ashley’s experimental report the sources were used cumulatively, because the experiments presented a revised version of the previous experiment. This need not always be so; in other experimental reports the sources may be independent of each other and may independently increase or decrease the plausibility of a certain hypothesis (see Kertész & Rákosi 2012).

- premises which certain plausible inferences constituting the argumentation structure of the experimental report are built on.
- (b) Through the cycles of the retrospective re-evaluation of information the result of real experiments may modify the plausibility of the result of the thought experiment included in the starting p-context.
 - (c) Thought experiments are constitutive parts of real experiments: the experimental design summarises the researcher's previous considerations about the workability of the experiment. Its task is to check whether the experimental procedure will be capable of producing reliable results.
 - (d) Thought experiments and real experiments are *closely related* and *have a similar structure*.
 - (e) Both real and thought experiments provide *plausible, but not certainly true, experimental/thought experimental data*.
 - (f) There are also important differences between thought experiments and real experiments.
 - (g) Therefore, they have different strengths and weak points. From this it follows that the *integration* of real and thought experiments as data sources within some research may considerably enhance the reliability of the hypotheses obtained.

The present chapter started with the remark that, while on the one hand, in contemporary linguistics there is a need to integrate different data sources, on the other hand, the methodology which should govern such an integration is missing. By obtaining (SP) with the help of the p-model, we have made an attempt to fill this gap with respect to the relationship between thought experiments and real experiments in pragmatics. Nevertheless, to reveal additional aspects of this relationship, numerous further case studies should be carried out.

Acknowledgements

Work on this chapter was supported by the MTA-DE Research Group for Theoretical Linguistics, the project OTKA K 77823 as well as the TÁMOP-4.2.2/B-10/1-2010-0024 project, which is co-financed by the European Union and the European Social Fund. Csilla Rákosi's work was also supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

- Austin, John L. 1962. *How to do things with words*. Oxford: Oxford University Press.
- Brown, James Robert & Fehige, Yiftach. 2011. Thought experiments. *Stanford Encyclopedia of Philosophy*. <<http://plato.stanford.edu/entries/thought-experiment/>> (September 23, 2013).
- Cohen, Philip R. & Levesque, Hector J. 1990. Rational interaction as the basis for communication. In *Intentions in communication*, Philip R. Cohen, Jerry L. Morgan & Martha E. Pollack (eds), 221-256. Cambridge, MA: MIT Press.
- Collins, Harry M. 1985. *Changing order: Replication and induction in scientific practice*. Beverly Hills & London: Sage.

- Gibbs, Raymond W. 1983. Do people always process the literal meaning of indirect requests? *Journal of Experimental Psychology: Learning, Memory & Cognition* 9: 524-533.
- Gooding, David. 1993. What is Experimental about Thought Experiment? In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1992, Volume Two: Symposia and Invited Papers (1992), 280-290. Michigan: Philosophy of Science Association.
- Grice, Paul. 1989a [1957]. Meaning. In Grice (1989c), 213-223.
- Grice, Paul. 1989b. Utterer's meaning and intentions. In Grice (1989c), 86-116.
- Grice, Paul. 1989c. *Studies in the way of words*. Cambridge, MA.: Harvard University Press.
- Holtgraves, Thomas & Ashley, Aaron. 2001. Comprehending illocutionary force. *Memory & Cognition* 29: 83-90.
- Kertész, András. 2010. On Stage Setting Thought Experiments in Cognitive Linguistics: A Case Study. In *Von Katastrophen, Zeichen und vom Ursprung der menschlichen Sprache*, Cornelia Stroh (Hrsg.), 7-30. Bochum: Brockmeyer.
- Kertész, András & Kiefer, Ferenc. 2013. From Thought Experiments to Real Experiments in Pragmatics. In *Perspectives on Philosophy and Pragmatics*, Alessandro Capone, Franco Lo Piparo & Marco Carapezza (eds), Berlin, Heidelberg & New York: Springer (in press).
- Kertész, András & Rákosi, Csilla. 2012. *Data and evidence in linguistics: A plausible argumentation model*. Cambridge: Cambridge University Press.
- Kertész, András & Rákosi, Csilla. This volume a. Introduction: The state of the art and the structure of the book.
- Kertész, András & Rákosi, Csilla. This volume b. The p-model of data and evidence in linguistics.
- Keysar, Boaz, Shen, Yeshayahu, Glucksberg, Sam & Horton, William S. 2000. Conventional language: How metaphorical is it? *Journal of Memory and Language* 43: 576-593.
- Kühne, Ulrich. 2005. *Die Methode des Gedankenexperiments*. Frankfurt am Main: Suhrkamp.
- Meggle, Georg & Ulkan, Maria. 1979. Kennst du das Land, wo die Zitronen blühen? (Zu Searles Kritik an Grice). *Papiere zur Linguistik* 20: 75-87.
- Meibauer, Jörg. 2012. Pragmatic evidence, context, and story design. An essay on recent developments in experimental pragmatics. *Language Sciences* 34(6) [= Kertész, András, Schwarz-Friesel, Monika & Consten, Manfred (eds), *Converging Data Types in Cognitive Linguistics*. Elsevier, Amsterdam etc.], 768-776.
- Moue, Aspasia S., Masavetas, Kyriakos A. & Karayianni, Haido. 2006. Tracing the developments of thought experiments in the philosophy of natural sciences. *Journal for the General Philosophy of Science* 37: 61-75.
- Nersessian, Nancy J. 1993. In the theoretician's laboratory: Thought experimenting 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.
- Nersessian, Nancy J. 1991. The Cognitive Sciences and the History of Science. In *Proceedings, joint HSS-SHOT conference on Critical Problems and Research Frontiers in History of Science and Technology, October 1991*, 92-116. Madison.
- Norton, John D. 1991. Thought experiments in Einstein's work. In *Thought Experiments in Science and Philosophy*, Tamara Horowitz & Gerard J. Massey (eds), 129-48. Savage, MD: Rowman and Littlefield.
- Norton, John D. 2002. On Thought Experiments: Is There More to the Argument? <<http://philsci-archive.pitt.edu/960/1/Norton.doc>>
- Noveck, Ira A. & Sperber, Dan. 2007. The why and how of experimental pragmatics. In *Pragmatics*, Noel Burton-Roberts (ed), 184-212. Basingstoke: Palgrave Macmillan.
- Noveck, Ira A. & Reboul, Anne. 2008. Experimental Pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Sciences* 12(11): 425-431.

- Rákosi, Csilla. 2011a. Metatheoretical reconstruction of psycholinguistic experiments. Part 1. *Sprachtheorie und germanistische Linguistik*, 21.1 [= Csatár, Péter (ed), Metapher, Metonymie und Argumentation / Metaphor, Metonymy and Argumentation 1. Münster: Nodus], 55-93.
- Rákosi, Csilla. 2011b. Metatheoretical reconstruction of psycholinguistic experiments. Part 2. *Sprachtheorie und germanistische Linguistik*, 21.2 [= Csatár, Péter (ed), Metapher, Metonymie und Argumentation / Metaphor, Metonymy and Argumentation 2. Münster: Nodus], 159-187.
- Rákosi, Csilla. 2012. The Fabulous Engine: Strengths and flaws of psycholinguistic experiments. *Language Sciences* 34(6) [= Kertész, András, Schwarz-Friesel, Monika & Consten, Manfred (eds), *Converging Data Types in Cognitive Linguistics*. Elsevier, Amsterdam etc.], 682-702.
- Schegloff, Emanuel A. 1991. Conversational analysis and socially shared cognition. In *Perspectives on socially shared cognition*, Lauren B. Resnick, John M. Levine & Stephanie D. Teasley (eds), 150-171. Washington DC: American Psychological Association.
- Searle, John R. 1969. *Speech acts. An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Searle, John R. & Vanderveken, Daniel. 1985. *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.
- Walton, Douglas, Reed, Chris & Macagno, Fabrizio. 2008. *Argumentation schemes*. Cambridge: Cambridge University Press.
- Wittgenstein, Ludwig. 1958 [1953]. *Philosophical investigations*. Oxford: Blackwell.