Proceedings of the Twenty-Ninth International
Florida Artificial Intelligence Research Society Conference

A Formal Model of Plausibility Monitoring in Language Comprehension

Maj-Britt Isberner
Department of Psychology
University of Kassel, Germany

Gabriele Kern-Isberner
Department of Computer Science
TU Dortmund, Germany

Abstract

Recent work in psychology provided evidence that plausibility monitoring is a routine component of language comprehension by showing that reactions of test persons were delayed when, e.g., a positive response was required for an implausible target word. These experimental results raise the crucial question of whether, and how, the role of plausibility assessments for the processes inherent to language comprehension can be made more precise. In this paper, we show that formal approaches to plausibility from the field of knowledge representation can explain the observed phenomena in a satisfactory way. In particular, we argue that the delays in response time are caused by belief revision processes which are necessary to overcome the mismatch between plausible context (or background resp. world) knowledge and implausible target words.

1 Introduction

In psychology, the question of whether linguistic information is routinely evaluated for truth or plausibility based on relevant world knowledge (*epistemic validation*, see, e.g., (Richter, Schroeder, and Wöhrmann 2009)) during language comprehension is still a point of contention. A widely accepted view is that epistemic validation is a strategic, optional process which is subsequent to language comprehension (Gilbert 1991; Gilbert, Tafarodi, and Malone 1993). Based on this idea, two-step models of comprehension and validation have been predominant which either assume that comprehension proceeds without any evaluative component (e.g., (Connell and Keane 2006)), or that the linguistic input is by default initially accepted as true and can only effortfully be unbelieved at a later point (e.g., (Gilbert, Tafarodi, and Malone 1993)). Thus, a common assumption of these two-step models is that readers need to actively question the plausibility of information to notice inconsistencies with their world knowledge. This would imply that it is possible for readers to comprehend linguistic information while ignoring whether or not it is plausible.

However, most modern theories of language comprehension agree that to understand a text, readers need to integrate text information with their knowledge about the world to construct a situation model of what the text is about (Johnson-Laird 1983; van Dijk and Kintsch 1983; Zwaan and Radvansky 1998). An important but generally overlooked implication of this assumption is that the process of constructing a situation model must be sensitive to the goodness of fit between incoming information and world knowledge. Therefore, (Isberner and Richter 2013) proposed that knowledge-based plausibility must be routinely monitored during language comprehension. They tested this assumption with a reaction time paradigm in which an assessment of plausibility was irrelevant or even detrimental to performance on the actual experimental task. In two experiments using different experimental tasks, they found interference of task-irrelevant plausibility with task performance, which constitutes evidence that readers cannot actually comprehend information without also assessing its consistency with their plausible beliefs about the world.

In this paper, we propose a formal model of what actually happens in the reader when he or she encounters plausible and implausible information of the kind used by (Isberner and Richter 2013) in their experiments, and discuss to what extent this model can account for their empirical findings. As a suitable framework for modelling plausible reasoning, we choose Spohn's ordinal conditional functions, *OCF* (Spohn 1988), and the approach of c-representations and c-revisions (Kern-Isberner 2004) because this combination is able to provide all methods necessary for a framework of plausible, inductive reasoning from background knowledge and iterated belief revision in the spirit of (Darwiche and Pearl 1997). C-representations allow for (inductive) nonmonotonic reasoning of a very high quality, meeting basically all standards which have been proposed for nonmonotonic logics so far (cf. (Kern-Isberner 2001; 2004)). Moreover, c-revisions generalize c-representations, so that we can take advantage of a seamless methodological framework for all reasoning activities that we consider in the experiments. This is important both from a formal and a psychological point of view because such a unifying theory adequately models the close link between uncertain reasoning and belief revision (cf., e.g., (Kern-Isberner 2001)) which has also been pointed out in the psychological literature (cf., e.g., (Politzer and Carles 2001)). However, we would like to emphasize that the focus here is on the formal reasoning activities themselves (inductive conditional reasoning, plausible reasoning, and iterated belief revision) as potential causes for observed delays. Conceivably, other unifying frameworks of plausible reasoning that provide all these reasoning activities in the same quality might work as well.

The basic idea is to simulate the test persons' reasoning by first setting up a knowledge base of conditionals which express the relevant beliefs for the situation under consideration in a task within an experiment. Instead of using some kind of plausibility distribution right away, we thereby aim at making plausible beliefs which form the relevant background knowledge that the test person uses in the tasks as explicit and transparent as possible. Then, an OCF-c-representation is built up which serves as an epistemic model of this background belief base, making the test person ready for responding to the respective task, which is presented in the form of two sentences. From this c-representation and the first sentence in the task, a situation model (or contextual epistemic state) is computed via revision techniques. Actually, we assume that this happens immediately when the person reads the first sentence. Then, when reading the second sentence, after which a response is expected, this second information is incorporated into the contextual epistemic state via a suitable revision operation (which may even be vacuous) before responding. Our claim is that this revision takes more or less time, depending on how compatible the second information is with the contextual epistemic state after the first information. Using one typical example from the experiments of (Isberner and Richter 2013), we will illustrate in detail what happens when the two pieces of information come in, and explain why response delays may occur.

This paper is organized as follows: In section 2, we fix some logical notations and recall how c-representations and c-revisions based on ordinal conditional functions (OCFs) can serve as a general model for plausible (inductive) reasoning. Section 3 describes briefly the details of the experiments of (Isberner and Richter 2013) and summarizes their results. Moreover, we set up the example that is used for illustrating our approach here. In section 4, we present our formal epistemic modelling for the experimental findings in (Isberner and Richter 2013). Section 5 concludes the paper by highlighting its main contributions and pointing out future research questions.

2 Reasoning and revision with OCFs

We build upon a propositional logical framework. Let $\mathcal{L}$ be a finitely generated propositional language, with atoms $a, b, c, \ldots$, and with formulas $A, B, C, \ldots$. For conciseness of notation, we will omit the logical *and*-connector, writing AB instead of $A \wedge B$, and overlining formulas will indicate negation, i.e. $\overline{A}$ means $\neg A$. Let Ω denote the set of possible worlds over $\mathcal{L}$; Ω will be taken here simply as the set of all propositional interpretations over $\mathcal{L}$. $\omega \models A$ means that the propositional formula $A \in \mathcal{L}$ holds in the possible world $\omega \in \Omega$; then ω is a *model* of A. As usual, let $\models$ also denote the classical entailment relation between propositions. By slight abuse of notation, we will use ω both for the model and the corresponding conjunction of all positive or negated atoms. The classical consequences of a set $\mathcal{S}$ of formulas are given by $\mathit{Cn}(\mathcal{S}) = \{B \in \mathcal{L} \mid \bigwedge \mathcal{S} \models B\}$.

Conditionals $(B|A)$ over $\mathcal{L}$, i.e., $A, B \in \mathcal{L}$, are meant to express uncertain, plausible rules "If A then plausibly B". The language of all conditionals over $\mathcal{L}$ will be denoted by $(\mathcal{L}|\mathcal{L})$. A conditional $(B|A)$ is *verified* by a world ω if $\omega \models AB$, and *falsified* if $\omega \models A\overline{B}$. A (finite) set $\mathcal{R} = \{(B_1|A_1), \ldots, (B_n|A_n)\} \subset (\mathcal{L}|\mathcal{L})$ expresses plausible beliefs of a human being or an agent, and is called a *(conditional) knowledge base*. We will use the terms knowledge and beliefs rather synonymously to denote propositions the agent is strongly convinced of, or deems to be most plausible. Knowledge bases $\mathcal{R}$ should be consistent in the sense that they should represent a coherent world view of the agent. This is certainly the case if it is possible to validate the plausibility of all conditionals of $\mathcal{R}$ within a formal epistemic framework. Such an epistemic framework can be set up via so-called ordinal conditional functions.

Ordinal conditional functions (*OCFs*, also called *ranking functions*) $\kappa : \Omega \to \mathbb{N} \cup \{\infty\}$ with $\kappa^{-1}(0) \neq \emptyset$, were introduced first by (Spohn 1988). They express degrees of plausibility of propositional formulas A by specifying degrees of disbeliefs (or implausibility) of their negations $\overline{A}$. More formally, we have $\kappa(A) := \min\{\kappa(\omega) \mid \omega \models A\}$, so that $\kappa(A \vee B) = \min\{\kappa(A), \kappa(B)\}$. Hence, due to $\kappa^{-1}(0) \neq \emptyset$, at least one of $\kappa(A), \kappa(\overline{A})$ must be 0, and altogether we have $\kappa(\top) = 0$ where $\top$ denotes a tautology. A proposition A is believed if $\kappa(\overline{A}) > 0$ (which implies particularly $\kappa(A) = 0$). Degrees of plausibility can also be assigned to conditionals by setting $\kappa(B|A) = \kappa(AB) - \kappa(A)$. A conditional $(B|A)$ is *accepted* in the epistemic state represented by κ, written as $\kappa \models (B|A)$, iff $\kappa(AB) < \kappa(A\overline{B})$, i.e. iff AB is more plausible than $A\overline{B}$. This can also be understood as a plausible, nonmonotonic inference: From A, we can plausibly derive B, in symbols $A \mid\!\sim_\kappa B$, iff $\kappa(AB) < \kappa(A\overline{B})$. In this way, OCFs can provide semantics for validating conditionals and plausible inference, and have become quite a popular model for (non-quantitative) epistemic states (Goldszmidt and Pearl 1996). The most plausible beliefs represented by an OCF κ are contained in the set $\mathit{Bel}\,(\kappa)$ which is the set of all formulas that are satisfied by all most plausible models, i.e., by all ω with $\kappa(\omega) = 0$. More formally, we have $\mathit{Bel}\,(\kappa) = \mathit{Cn}(\vee_{\kappa(\omega)=0}\omega)$. OCF-rankings can be understood as logarithmic probabilities (Goldszmidt and Pearl 1996), so there are lots of analogies between OCFs and probability functions. In particular, the *uniform OCF* κ_u assigns the same rank to each world: $\kappa_u(\omega) = 0$ for all $\omega \in \Omega$. Note that OCFs treat (plausible) propositions A in the same way as the conditional $(A|\top)$, so we consider only conditionals as elements of our knowledge bases but keep in mind that knowledge bases can also contain plausible propositions.

Given a knowledge base $\mathcal{R}$ which usually expresses only partial knowledge about the world, a crucial task is to find an epistemic state that validates $\mathcal{R}$ and completes it with plausible inferences that can be drawn from $\mathcal{R}$. This process of completing a knowledge base towards an epistemic state is often called *inductive reasoning*. In our framework, this means that we have to compute an OCF κ that accepts all conditionals in $\mathcal{R}$ and can then be used for plausible reasoning. We will use here the approach of c-representations that allows for an easy generalization to also handle revision tasks (Kern-Isberner 2001; 2004).

A *c-representation* of a knowledge base $\mathcal{R} = \{(B_1|A_1), \ldots, (B_n|A_n)\}$ is an OCF κ of the form

$$\kappa_{\mathcal{R}}(\omega) = \sum_{\omega \models A_i\overline{B_i}} \kappa_i^- \tag{1}$$

with non-negative integers κ_i^- that are chosen in such a way as to ensure that $\kappa \models \mathcal{R}$, i.e.,

$$\kappa_j^- > \min_{\omega \models A_j B_j} \sum_{\substack{i \neq j \\ \omega \models A_i\overline{B_i}}} \kappa_i^- - \min_{\omega \models A_j\overline{B_j}} \sum_{\substack{i \neq j \\ \omega \models A_i\overline{B_i}}} \kappa_i^- . \tag{2}$$

Having one epistemic state is not enough – new information comes in, and the agent (or the human) has to incorporate this information into her epistemic state. So, the epistemic state has to be changed, i.e., we have to perform a belief change operation (usually denoted with the symbol $*$) to adapt the agent's beliefs to the current state of the world. Belief revision theory provides lots of approaches to tackle this problem (for a recent overview, see (Fermé and Hansson 2011)). In this paper, we make use of c-revisions (Kern-Isberner 2001; 2004) as a powerful approach to handle advanced belief change tasks that we need here. More precisely, c-revisions are able to solve the following problem: Given a prior OCF κ and a set $\mathcal{R}$ of conditionals that represent new information, compute a posterior $\kappa^* = \kappa * \mathcal{R}$ that accepts $\mathcal{R}$ and still uses as much of the information of the prior κ as possible. C-revisions are built in a way that is very similar to c-representations; actually, c-representations arise from c-revisions when a uniform epistemic state κ_u is revised by $\mathcal{R}$. For this paper, we only need c-revisions in a simpler form because the new information will only be one plausible proposition, that is, we only have to find a revision of an epistemic prior κ by a proposition A. For this case, a handy and unique form of c-revisions can be used (for more details, please see (Kern-Isberner and Huvermann 2015)):

$$\kappa * A(\omega) = \begin{cases} \kappa(\omega) - \kappa(A) & \text{, if } \omega \models A \\ \kappa(\omega) + \max\{0, -\kappa(\overline{A}) + 1\}, & \text{if } \omega \models \overline{A} \end{cases} \tag{3}$$

If already $\kappa \models A$, then $\kappa * A = \kappa$; otherwise, A-worlds are shifted downwards by $\kappa(A)$, and $\overline{A}$-worlds are shifted upwards by 1. In any case, after the revision we have $\kappa * A(\overline{A}) > 0$, as required. $\kappa * A$ is a kind of conditioning of κ by A. In (Kern-Isberner 2004; Kern-Isberner and Huvermann 2015), technical details for generalizing this to several new pieces of (more complex) information can be found.

After having laid the formal grounds for our approach to explaining monitoring in language comprehension, we explain how language comprehension is evaluated in (Isberner and Richter 2013) in more detail.

3 Language comprehension in psychology

In the experiments by Isberner and Richter (Isberner and Richter 2013), participants read sentence pairs describing everyday situations that were either plausible or implausible with regard to general world knowledge. These sentence pairs were presented word by word on a computer screen (300ms/word) and their plausibility always hinged on the last word of the sentence pair (target word). 300 ms after the target word appeared, participants were prompted to perform a task on the target word. The task was either an orthographic task in which participants were asked to indicate whether the target word was spelled correctly or not (Experiment 1), or a color judgment task in which participants judged whether or not the target word changed color (Experiment 2). Thus, participants in both tasks were required to provide positive (yes) and negative (no) responses unrelated to the plausibility of the preceding sentence pair. (Isberner and Richter 2013) predicted that if plausibility monitoring is routine, responses should be delayed when the required response is incongruent with the presumed outcome of the routine (but task-irrelevant) plausibility assessment compared to when it is congruent. The results of both experiments confirmed this prediction but only for positive responses, which were significantly slower for target words that made the described situation implausible than for target words that made it plausible, while negative reponses were either also slower for implausible target words (Experiment 1) or not affected (Experiment 2). We will focus on Experiment 1 in the following.

Condition	Item version
Plausible, predictable word	Frank has a broken pipe. He calls the plumber.
Implausible, predictable word	Frank has a broken leg. He calls the plumber.
Plausible, unpredictable word	Frank has a broken pipe. He calls the tradesman.
Implausible, unpredictable word	Frank has a broken leg. He calls the tradesman.

Table 1: Sentence pairs from Experiment 1 by (Isberner and Richter 2013)

(Isberner and Richter 2013) also tried to rule out that their results could be explained by predictability rather than plausibility (plausible target words are usually more predictable than implausible target words) by using target words that were either predictable or unpredictable, where (un)predictability is always assessed with respect to the plausible context. Predictability had been ascertained experimentally before. In the context of knowledge representation, predictability could be interpreted as specificity, or informativeness. Although unpredictable target words were generally responded to more slowly, the overall pattern of a delay of positive responses to implausible as compared to plausible target words did not significantly differ from the pattern for predictable target words, which supports the prediction that this delay is due to plausibility rather than predictability.

In (Isberner and Richter 2013), a within-items manipulation of plausibility and predictability is used, meaning that for each possible combination of both variables (i.e., for each condition), a different version of the same item was constructed. Each participant saw only one version of each item but the same number of items in each condition, such that across participants, all versions of each item were used.

We chose the items from Table 1 as a typical example of the stimuli in (Isberner and Richter 2013). Each pair of sentences makes up one task. The item version shows the sentences while the condition explains the precise combination of plausibility and predictability used for the respective task. Plausibility and predictability refer to the last word of each sentence pair (given the context), which is crucial for the task and hence underlined.

Table 2 shows the average response times for positive and negative answers with respect to each combination of plausibility and predictability in Experiment 1. It is clearly seen that in any case, responses were significantly slower for implausible words.

Condition	*Plausible*		*Implausible*	
	Response time		Response time	
	M	(SD)	M	(SD)
Predictable				
Positive response	953	(280)	1034	(333)
Negative response	873	(261)	927	(331)
Unpredictable				
Positive response	1026	(293)	1171	(371)
Negative response	995	(317)	1009	(310)

Table 2: Response times (means (M) and standard deviations (SD) by experimental condition) of Experiment 1 in (Isberner and Richter 2013). Means and standard deviations are based on participants as units of observations.

In the next section, we show how to explain the findings by (Isberner and Richter 2013) with the help of the reasoning and revision techniques described in section 2.

4 OCFs as an epistemic model for language comprehension

The insights gained from the experiments described in section 3 suggest that human beings perform some mental action that take some milliseconds when being faced with implausible statements, but prima facie it is not clear what exactly happens in their minds, and why plausible statements are processed more smoothly. In this section, we underpin the findings from (Isberner and Richter 2013) by setting up an epistemic model for a typical example used in the experiments with the help of OCFs. By the formal means of inductive reasoning and belief revision we are able to simulate the information processing within a test person's mind and to explain the observations from (Isberner and Richter 2013). A key feature of our approach is that we will make background knowledge explicit, since in all examples from (Isberner and Richter 2013), test persons are expected to use commonsense knowledge to validate the statements.

In all examples, the course of the task is the same, and is formally modelled as follows: In the beginning, before receiving the first information, the epistemic state of the test person is given by some OCF κ which serves as a basis for validation for all four sentences. This κ must reflect relevant background knowledge in some way. Then, the first information A arrives, triggering an adaptation of κ to A to set up the epistemic context (situation model) for the task. In our approach, this is modelled by a belief revision operation $\kappa{*}A$ which we realize by c-revisions. Then the second information B arrives, and the crucial question for evaluating the plausibility of B is: What is the formal-logical relationship between $\kappa{*}A$ and B? If $\kappa{*}A \models B$, then B is plausible in the context of $\kappa * A$, but if $\kappa * A \models \neg B$, then there is a conflict between the new information B and what the test person's current epistemic context validates as plausible. Solving this conflict, or even merely deciding to ignore the conflict, takes time and causes the observed delay.

We will explain this in more detail with the example from Table 1. First, we have to take care of modelling relevant background knowledge. We use the following logical variables for this:

P	having a broken pipe	L	having a broken leg
U	calling the plumber	T	calling the tradesman
D	calling the doctor		

For each variable V, v means "V is true", and $\overline{v}$ means "V is false", so e.g., p symbolizes the sentence "Frank has a broken pipe." We specify relevant background knowledge by the following knowledge base $\mathcal{R} = \{(u|p), (d|l), (t|u), (\overline{t}|d)\}$:

$(u\|p)$	If one has a broken pipe, one usually calls the plumber.
$(d\|l)$	If one has a broken leg, one usually calls the doctor.
$(t\|u)$	A plumber is (usually) a tradesman.
$(\overline{t}\|d)$	A doctor is usually not a tradesman.

Applying the technique of c-representations and choosing minimal parameters κ_i^-, we obtain the OCF κ shown in Table 3. The calculations for setting up κ are straightforward using equations (1) and (2), so we explain them with just some examples.

For a c-representation of $\mathcal{R}$, we need four parameters $\kappa_1^-, \kappa_2^-, \kappa_3^-, \kappa_4^-$, each κ_i^- being associated with the i-th conditional in $\mathcal{R}$, e.g., κ_2^- is associated with the second conditional $(d|l)$. Since there are no direct conflicts between the conditionals in $\mathcal{R}$, we obtain for all four parameters $\kappa_i^- > 0$ from (2), so we can choose them minimally by setting $\kappa_i^- = 1$ for $i = 1, \ldots, 4$. This means that for all worlds ω, we just have to count how many conditionals in $\mathcal{R}$ are falsified by ω, according to (1). For example, the world $uptld$ (in which Frank has a broken pipe and a broken leg, and he calls someone who is a plumber, a doctor, and a tradesman[1]) falsifies only the fourth conditional $(\overline{t}|d)$ and hence has the κ-rank 1. Analogously, the world $up\overline{t}l\overline{d}$ (in which Frank has a broken pipe and a broken leg, and he calls someone who is a plumber, but neither a tradesman nor a doctor) falsifies $(d|l)$

[1] Note that in a propositional framework, we cannot model statements on different objects, so in order to keep the modelling consistent, it must always be Frank who has a broken pipe or a broken leg, and it must always be some other person who can be a plumber, a doctor, or a tradesman.

ω	κ	$\kappa * p$	$\kappa * l$	$\kappa * \overline{u}t$	ω	κ	$\kappa * p$	$\kappa * l$	$\kappa * \overline{u}t$
$uptld$	1	1	1	2	$\overline{u}ptld$	2	2	2	1
$uptl\overline{d}$	1	1	1	2	$\overline{u}ptl\overline{d}$	2	2	2	1
$upt\overline{l}d$	1	1	2	2	$\overline{u}pt\overline{l}d$	2	2	3	1
$upt\overline{l}\,\overline{d}$	0	0	1	1	$\overline{u}pt\overline{l}\,\overline{d}$	1	1	2	0
$up\overline{t}ld$	1	1	1	2	$\overline{u}p\overline{t}ld$	1	1	1	2
$up\overline{t}l\overline{d}$	2	2	2	3	$\overline{u}p\overline{t}l\overline{d}$	2	2	2	3
$up\overline{t}\,\overline{l}d$	1	1	2	2	$\overline{u}p\overline{t}\,\overline{l}d$	1	1	2	2
$up\overline{t}\overline{l}\,\overline{d}$	1	1	2	2	$\overline{u}p\overline{t}\,\overline{l}\,\overline{d}$	1	1	2	2
$u\overline{p}tld$	1	2	1	3	$\overline{u}\,\overline{p}tld$	1	2	1	1
$u\overline{p}tl\overline{d}$	1	2	1	3	$\overline{u}\,\overline{p}tl\overline{d}$	1	2	1	1
$u\overline{p}t\overline{l}d$	1	2	2	3	$\overline{u}\,\overline{p}t\overline{l}d$	1	2	2	1
$u\overline{p}t\overline{l}\,\overline{d}$	0	1	1	2	$\overline{u}\,\overline{p}t\overline{l}\,\overline{d}$	0	1	1	0
$u\overline{p}\overline{t}ld$	1	2	1	3	$\overline{u}\,\overline{p}\overline{t}ld$	0	1	0	2
$u\overline{p}\overline{t}l\overline{d}$	2	3	2	4	$\overline{u}\,\overline{p}\overline{t}l\overline{d}$	1	2	1	3
$u\overline{p}\overline{t}\,\overline{l}d$	1	2	2	3	$\overline{u}\,\overline{p}\overline{t}\,\overline{l}d$	0	1	1	2
$u\overline{p}\overline{t}\,\overline{l}\,\overline{d}$	1	2	2	3	$\overline{u}\,\overline{p}\overline{t}\,\overline{l}\,\overline{d}$	0	1	1	2

Table 3: Prior κ and revised OCFs $\kappa * p$, $\kappa * l$, and $\kappa * \overline{u}t$ for the example from Table 1

and $(t|u)$ and hence is assigned the rank $\kappa_2^- + \kappa_3^- = 2$. Generally, the more conditionals from $\mathcal{R}$ ω falsifies, the higher is its κ-rank, and the less plausible ω is. Exactly six worlds do not falsify any conditional in $\mathcal{R}$ and thus have κ-rank 0: These are the models $upt\overline{l}\,\overline{d}, u\overline{p}t\overline{l}\,\overline{d}, \overline{u}\,\overline{p}t\overline{l}\,\overline{d}, \overline{u}\,\overline{p}\overline{t}ld, \overline{u}\,\overline{p}\overline{t}\,\overline{l}d$ and $\overline{u}\,\overline{p}\overline{t}\,\overline{l}\,\overline{d}$, so in her initial epistemic state, the agent accepts all conditionals in $\mathcal{R}$ but is indifferent with respect to all logical variables P, L, U, T, D, that is, she believes neither the positive nor the negative form of each variable.

Then the first information comes in: "Frank has a broken pipe/leg", i.e., the test person comes to know p resp. l and has to incorporate this information into κ. We compute $\kappa * p$ resp. $\kappa * l$ which are also shown in Table 3. Let us first consider $\kappa * p$ which is computed via (3) by shifting $\overline{p}$-worlds upwards by 1. Since there is only one world ω with $\kappa * p(\omega) = 0$, namely $\omega = upt\overline{l}\,\overline{d}$, we have $\kappa * p \models upt\overline{l}\,\overline{d}$ – after reading "Frank has a broken pipe", the agent believes that Frank has a broken pipe and that he calls the plumber (who is a tradesman), but she does not believe that he has a broken leg, nor that he calls the doctor. So, when she then reads that "He calls the plumber", this fits her beliefs perfectly. Therefore, a revision $\kappa * p$ by the new information u is effortless, we have $(\kappa * p) * u = \kappa * p$, and thus does not cause any delay. However, if the agent first comes to know "Frank has a broken leg", her revision $\kappa * l$ yields belief in $\overline{u}\,\overline{p}\overline{t}ld$ – now she believes that Frank has a broken leg and calls the doctor, but also that Frank does not have a broken pipe and in particular, that he does not call the plumber (nor a tradesman). So, the next information "He calls the plumber" is contradictory to what she believes, and the adaptation of $\kappa * l$ to u needs a true revision to solve the conflict: $(\kappa * l) * u \neq (\kappa * l)$. For the same reason, the sentences *Frank has a broken leg – He calls the tradesman* cause a confusion of the test person because after learning *Frank has a broken leg*, she believes $\overline{u}\,\overline{p}\overline{t}ld$, so *tradesman* is implausible.

The example *Frank has a broken pipe – He calls the tradesman* is more intricate. In the experiments, Isberner and Richter (Isberner and Richter 2013) noticed a slight delay here in any case. This cannot be explained straightforwardly by our modelling since when learning *Frank has a broken pipe*, also *tradesman* is plausible ($\kappa * p \models upt\overline{l}\,\overline{d}$). A first explanation for this effect can be given by looking closer at the knowledge base $\mathcal{R}$: Here, the conditional $(u|p)$ establishes an immediate connection between p and u, while t is entailed from p only via a transitive chaining of the conditionals $(u|p)$ and $(t|u)$. This would imply that the agent does not reason from the full epistemic state $\kappa * p$ in any case but takes the knowledge base as a more compact representation of her beliefs. Only in cases where she is not able to derive an answer directly from the knowledge base, she initiates the more complex reasoning process of computing $\kappa * p$. Note that (naive) transitive chaining is not allowed in general because other conditionals might interfere. But in the case considered in this particular example, transitive chaining of $(u|p)$ and $(t|u)$ would be allowed since $\kappa \models (t|p)$ because of $\kappa(pt) = 0 < 1 = \kappa(p\overline{t})$.

Another explanation could be as follows: Reading t leaves the test person with the options ut or $\overline{u}t$. She may assume that the information given to her in the test is as specific as possible (which would be u, or ut). That is, after reading t, she might wonder whether actually $\overline{u}t$ is meant. But $\overline{u}t$ is as implausible as d in the context of $\kappa * p$: $\kappa * p(\overline{u}t) = 1$ while $\kappa * p(\overline{\overline{u}t}) = 0$. Because of $\kappa * p(ut) = 0 < \kappa * p(\overline{u}t) = 1$, the test person accepts the conditional $(u|t)$ in the epistemic state $\kappa * p$ – *if a tradesman is called it is plausibly a plumber*. The option $\overline{u}t$, which might appear realistic, would not only violate the current beliefs of the test person, its incorporation $(\kappa * p) * \overline{u}t$ which is also shown in Table 3 even casts doubt on p being true or not: We have $\mathit{Bel}\,((\kappa * p) * \overline{u}t) = \mathit{Cn}(\overline{u}pt\overline{l}\,\overline{d} \vee \overline{u}\,\overline{p}t\overline{l}\,\overline{d}) = \mathit{Cn}(\overline{u}t\overline{l}\,\overline{d})$, that is, the test person would be uncertain whether p still holds. Anticipating these problems, the test person might adhere to the plausible option ut, but even deliberating about this costs time and may cause delays.

The symmetric effect of positive vs. negative responses observed (but not expected) by (Isberner and Richter 2013) in the experiments – both are significantly delayed – appears to be completely reasonable on the basis of our model because the revision processes depend only on the (logical) incompatibility between contextual knowledge and new information, not on the polarity of the response.

5 Conclusion and future work

In this paper, we set up a formal model of plausible reasoning and belief revision that helps explain findings in psychological experiments for language comprehension. We simulated the test persons' epistemic processes when reading the texts given in the experiments by first reasoning inductively from background knowledge bases, then building up a situation model (contextual epistemic model) from this, and afterwards evaluating and incorporating new information with respect to this epistemic context. We argue that the observed delays in responding to given tasks are caused by revision processes that are necessary to overcome incompatibilities between plausible context knowledge and obtained information.

We consider our paper as providing the first steps towards underpinning experimental findings in language comprehension with formal models from knowledge representation. Both disciplines may benefit from that: Actually, the experiments in (Isberner and Richter 2013) make use of a variety of different kinds of knowledge, e.g., declarative knowledge from experience, causal knowledge, and normative knowledge. We will categorize these kinds of knowledge and reinvestigate whether significant differences can be found when dealing with different kinds of knowledge. If so, the formal model can be fine-tuned and optimized to allow for distinguishing between these different reasoning modes. Moreover, the design of the experiments can also be modified to investigate more refined research hypotheses: We showed that the validation process checks the second information B for plausibility with respect to the epistemic context $\kappa * A$ and found that, if $\kappa * A \models B$, then the answer was quick, while if $\kappa * A \models \neg B$, then there was a delay. However, there is a third option, namely, $\kappa * A$ might be undecided with respect to B, i.e., it neither accepts B nor $\neg B$. We might expect that the answer would be somewhat delayed, but this has to be validated. From the viewpoint of language comprehension, an interesting question would be how the processes of routine monitoring of plausibility and the reasoning for responding are intertwined: In the experiments, a negative response was on average quicker than a positive one. This suggests that the processes might be performed at least partly in parallel, but further investigations are necessary here. We will also investigate whether such a hypothesized evaluation in parallel might also explain the asymmetry between positive vs. negative responses in the Experiment 2 of (Isberner and Richter 2013).

A research question that would be interesting for both disciplines is whether general differences between using explicit knowledge in the knowledge base and implicit knowledge derived by some epistemic processes can be observed. Basically, we abstracted from this effect here since we assumed that the situation model for evaluating the second sentence in the tasks is built up immediately. But the aspect of predictability may be related to that, as we pointed out in section 4. Moreover, our formal model that is based on degrees of plausibility might also be relevant to provide a logical environment to explain results from psychological experiments concerning uncertain reasoning and belief revision, as in (Politzer and Carles 2001)[2].

Our formal model to simulate human plausible reasoning can serve as a very helpful bridge between the AI discipline of knowledge representation and psychological research on language comprehension and general human reasoning. Crucially, our model is neither restricted to classical logics nor does it make use of probabilities but is located in the wide area of qualitative default logics between these two extremes. Therefore, we also consider our work as a proof of concept that suitable normative models for psychological experiments may be built from axiomatic or inductive approaches to plausible reasoning.

[2] We are grateful to an anonymous referee for pointing out this paper to us.

References

Connell, L., and Keane, M. T. 2006. A model of plausibility. *Cognitive Science* 95–120.

Darwiche, A., and Pearl, J. 1997. On the logic of iterated belief revision. *Artificial Intelligence* 89:1–29.

Fermé, E., and Hansson, S. 2011. AGM 25 years – twenty-five years of research in belief change. *Journal of Philosophical Logic* 40:295–331.

Gilbert, D. T.; Tafarodi, R. W.; and Malone, P. S. 1993. You can't not believe everything you read. *Journal of Personality and Social Psychology* 221–233.

Gilbert, D. T. 1991. How mental systems believe. *American Psychologist* 107–119.

Goldszmidt, M., and Pearl, J. 1996. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence* 84:57–112.

Isberner, M.-B., and Richter, T. 2013. Can readers ignore plausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica* 142:15–22.

Johnson-Laird, P. N. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge: Harvard University Press.

Kern-Isberner, G., and Huvermann, D. 2015. Multiple iterated belief revision without independence. In Russell, I., and Eberle, W., eds., *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, FLAIRS-28*. AAAI Press.

Kern-Isberner, G. 2001. *Conditionals in nonmonotonic reasoning and belief revision.* Springer, Lecture Notes in Artificial Intelligence LNAI 2087.

Kern-Isberner, G. 2004. A thorough axiomatization of a principle of conditional preservation in belief revision. *Annals of Mathematics and Artificial Intelligence* 40(1-2):127–164.

Politzer, G., and Carles, L. 2001. Belief revision and uncertain revision. *Thinking and Reasoning* 7(3):217–234.

Richter, T.; Schroeder, S.; and Wöhrmann, B. 2009. You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology* 538–558.

Spohn, W. 1988. Ordinal conditional functions: a dynamic theory of epistemic states. In Harper, W., and Skyrms, B., eds., *Causation in Decision, Belief Change, and Statistics, II.* Kluwer Academic Publishers. 105–134.

van Dijk, T. A., and Kintsch, W. 1983. *Strategies of discourse comprehension.* New York: Academic Press.

Zwaan, R. A., and Radvansky, G. A. 1998. Situation models in language comprehension and memory. *Psychological Bulletin* 162–185.