The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Rational Inference Patterns Based on Conditional Logic

Christian Eichhorn, Gabriele Kern-Isberner
Chair 1 of Computer Science
TU Dortmund University, Dortmund, Germany
christian.eichhorn@tu-dortmund.de
gabriele.kern-isberner@cs.tu-dortmund.de

Marco Ragni
Cognitive Computation Lab
University of Freiburg, Freiburg, Germany
ragni@informatik.uni-freiburg.de

Abstract

Conditional information is an integral part of representation and inference processes of causal relationships, temporal events, and even the deliberation about impossible scenarios of cognitive agents. For formalizing these inferences, a proper formal representation is needed. Psychological studies indicate that classical, monotonic logic is not the approriate model for capturing human reasoning: There are cases where the participants systematically deviate from classically valid answers, while in other cases they even endorse logically invalid ones. Many analyses covered the independent analysis of individual inference rules applied by human reasoners. In this paper we define *inference patterns* as a formalization of the joint usage or avoidance of these rules. Considering patterns instead of single inferences opens the way for categorizing inference studies with regard to their qualitative results. We apply plausibility relations which provide basic formal models for many theories of conditionals, nonmonotonic reasoning, and belief revision to asses the rationality of the patterns and thus the individual inferences drawn in the study. By this replacement of classical logic with formalisms most suitable for conditionals, we shift the basis of judging rationality from compatibility with classical entailment to consistency in a logic of conditionals. Using inductive reasoning on the plausibility relations we reverse engineer conditional knowledge bases as explanatory model for and formalization of the background knowledge of the participants. In this way the conditional knowledge bases derived from the inference patterns provide an explanation for the outcome of the study that generated the inference pattern.

1 Introduction

For a conditional statement "If A then B" and a respective fact, four inference patterns can be defined (not all are logically valid though): *Modus Ponens*, that from the statement and A it follows that B, and *Modus Tollens*, that from the statement and $\neg B$ it follows that $\neg A$, and the classically invalid rules *Affirmation of the Consequent*, stating that from the statement and B it follows that A and, finally, *Denial of the Antecedent*, stating that from the statement and $\neg A$ it follows that $\neg B$. Established and accepted studies show that all these inferences are indeed drawn by human reasoners, if presented with the "right" stimulus material. In the discussion of the results, these studies usually analyze the usage of each inference rule by the majority of the participants, separately.

In this paper, we consider the application or neglect of the our rules as an *inference pattern*. These patterns constitute a descriptive model that can explain the answers of an agent in solving conditional inference problems. With these inference patterns we propose a formal model for the studies on the assumption that the majority of the participants in these studies answered in a rational fashion. In this, the test cases for this formal basis are inference studies regarding human reasoning about a rule, rather than the participants of the studies themselves. Our proposed model is based upon conditional logic, that is, it is substantially different from classical as well as probabilistic logic, but combines interesting aspects of both. We assess the rationality of an inference pattern and by this the rationality of the inferences drawn based on the plausibility semantics of Ordinal Conditional Functions (OCF) (Spohn 1988). This shifts the assessment of rationality from rigidly following logical inference rules to satisfiability of plausibility constraints in semi-quantitative plausibility models situated between classical logic and probabilities.

For rational inference patterns, we use the inductive approach of c-representations (Kern-Isberner 2001) to algorithmically reverse engineer a conditional knowledge base to which the OCF that satisfies the inference pattern is admissible. In this way we ensure that these knowledge bases yield ranking models that are compatible with the inference patterns. We furthermore argue that this means that since each inductive ranking model of the knowledge base brings forth the inferences drawn in the study, the knowledge bases are hypotheses for the formalization of suitable background knowledge the participants may have activated when giving their answers in the studies. This means that the inductively generated knowledge bases serve as explanatory conditional models for the results of the considered inference study.

The paper is organized as follows: In the next section we recall relevant psychological findings regarding human conditional reasoning, from which two studies are to be used as running examples to illustrate the approach throughout the rest of the paper. We briefly define the formal preliminaries and notations necessary for the paper in Section 3. This

is followed by a brief introduction into plausible reasoning in general, in particular Ordinal Conditional Functions and the inductive approach of c-representations in Section 4. In Section 5 we propose global inference patterns and the constraints they impose on the plausibility relations, and in Section 6 we apply them to explain human inferences via an algorithmic approach. Section 7 sums up the findings and concludes the article.

2 Human Reasoning with Conditionals

Conditional reasoning is reasoning about statements of the form "if A then C", where A is an antecedent and C is a consequent. Psychologists have tested human inferences with such conditionals. Some results (Wason 1968) show a deviation of human conditional reasoning from the classical, bivalent truth table of the material implication. Explanations included a trivalent evaluation of the reasoner using defective (Wason 1968) or DeFinetti truth tables (Baratgin, Over, and Politzer 2013) which hints to an underlying conditional logic.

From the perspective of commonsense reasoning we briefly recall two core examples relevant for human reasoning: The first is the *suppression task*, demonstrating not only the use of non-logical rules, but also the suppression of classical inference rules in their conclusions(Byrne 1989). This paper started nonmonotonicity research in psychology and cognitive sciences. Participants have been given the rule "If Lisa has an essay to write, then she will study late in the library". Given the fact "Lisa has an essay to write", nearly all of the participants (96 %) concluded "She will study late in the library". But of participants who additionally received the rule "If the library is open, Lisa will study late in the library", only 38 % concluded "She will study late in the library" from the fact, in accordance with Modus Ponens (MP). The same difference (92 % to 33 %) could be found for the same task in a formulation matching Modus Tollens (MT), not concluding "Lisa does not have an essay to write" from the fact "She does not study late in the library". Also, the majority (63 %) of participants concluded "Lisa has an essay to write" from the fact "She studies late in the library", thus applying the rule Affirmation of the Consequent (AC) and "She will not study late in the library" from the fact "Lisa does not have an essay to write", applying Denial of the Antecedence (DA). These findings have been reproduced with different stimulus material and different aspects, see, for instance (Bonnefon and Hilton 2004; De Neys, Schaeken, and D'Ydewalle 2003; Politzer 2005) Recent work suggests that this retraction of classically valid inferences triggers a cautious rule approach and can be modeled by some nonmonotonic reasoning approaches (Ragni, Eichhorn, and Kern-Isberner 2016).

The second property of human reasoning that we focus on in this paper is the skill to reason counterfactually, that is, to reason hypothetically about facts that do not hold or did not hold in the past. For instance, the statement "If Oswald had not shot Kennedy, then someone else would have" presupposes what could have happened otherwise. This line of reasoning allows, among others, the reasoner to avoid undesirable outcomes of future events. Reasoning processes of factual and counterfactual conditionals are distinct and cognitive mental processes are different (Byrne and Egan 2004; Byrne and Tasso 1999; Egan and Byrne 2012; Egan, García-Madruga, and Byrne 2009; Frosch and Byrne 2012; Quelhas and Byrne 2003; Thompson and Byrne 2002). In (Thompson and Byrne 2002, Experiment 2), participants were asked to reason from the two conditionals: "If the car was out of gas, then it stalled." (factual) and "If the car had been out of gas, then it would have stalled." (counterfactual). Given the facts that "The car was not out of gas." or "The car did not stall.", the two negative inferences, DA and MT, had a much higher endorsement percentages in the counterfactual than in the factual case (50% and 30% for DA; and 85% and 68% for MT, respectively), while MP and AC did not change. Another study from (Byrne and Tasso 1999, Experiment 3) supports this finding with endorsement percentages for counterfactual and factual conditionals of 66% and 42% for MT; and 59% and 39% for DA, respectively. The results suggest that participants considered two alternatives when they encountered such counterfactual arguments, the fact and the supposed "fact" (also known as the "presupposed factual reality" and the "counterfactual conjecture"). (Thompson and Byrne 2002) argues that the construction of the alternative mental model of the supposed "fact" (the two "not" facts: not-(out of gas) and not-stall) in the counterfactual case facilitates the MT and DA inferences.

Note that these are just two examples for effects that can be found in human commonsense reasoning. Other examples include, but are not limited to, reasoning in different thematic environments (Griggs and Cox 1982), like, for instance, regarding social rules, or legal reasoning (Castañeda and Knauff 2016), where, in addition to the legal texts, emotions play a role in the conclusions of the reasoners. The formalism presented in this paper is capable of capturing any result of psychological studies regarding human conditional reasoning. We selected the two examples presented in this section to illustrate the presented formalism because of their prominence in reasoning research, only.

3 Formal Preliminaries

Our formal modeling is based on propositional logic with a language set up from a finite set of propositional atoms $\Sigma = \{V_1, \ldots, V_m\}$ which can be interpreted to be $true$ (v_i) or $false$ ($\overline{v}_i$). The propositional language $\mathfrak{L}$ is composed from Σ with the logical connectives *and* ($\wedge$), *or* ($\vee$), and *not* ($\neg$), as usual. We may leave out the symbol $\wedge$ and write conjunction by juxtaposition to obtain shorter formulas where no risk of confusion exists, and abbreviate negation ($\neg A$) by ($\overline{A}$). The set of possible worlds over Σ will be called Ω, we often use the 1-1 association between worlds and complete conjunctions, that is, conjunctions of literals $\dot{v}_i \in \{v_i, \overline{v}_i\}$ where every variable $V_i \in \Sigma$ appears exactly once. A formula $A \in \mathfrak{L}$ is evaluated under a world ω according to the classical logical rules, that is, $[\![A]\!]_\omega = true$ if and only if $\omega \models A$ if and only if $\omega \in Mod(A)$, that is, ω is an element of the classical models $Mod(A)$ of A. The set of classical consequences of a set of formulas $\mathcal{A} \subseteq \mathfrak{L}$ is $Cn(\mathcal{A}) = \{B | \mathcal{A} \models B\}$. The deductively closed set of formulas which has exactly a subset $\mathcal{W} \subseteq \Omega$ as models is

called the *formal theory* of $\mathcal{W}$ and defined as $Th(\mathcal{W}) = \{A \in \mathfrak{L} \mid \omega \models A \text{ for all } \omega \in \mathcal{W}\}$. The material implication "From A it (always) follows that B" is, as usual, equivalent to $\overline{A} \vee B$ and written as $A \Rightarrow B$.

We introduce the binary operator $|$ to obtain the set $(\mathfrak{L}|\mathfrak{L})$ of *conditionals* written as $(B|A)$. Conditionals are three-valued logical entities with the evaluation (DeFinetti 1974)

$$[\![(B|A)]\!]_\omega = \begin{cases} true & \text{iff } \omega \models AB \text{ (verification)} \\ false & \text{iff } \omega \models A\overline{B} \text{ (falsification)} \\ undefined & \text{iff } \omega \models \overline{A} \quad \text{(neutrality).} \end{cases}$$

For a conditional $(B|A)$, $(\overline{B}|A)$ is the *(strict) negation* of the conditional, and $(A|B)$ is its *inverse (form)*.

A *(conditional) knowledge base* is a finite set of conditionals $\Delta = \{(B_1|A_1), \ldots, (B_n|A_n)\} \subseteq (\mathfrak{L} \mid \mathfrak{L})$. To give appropriate semantics to conditionals and knowledge bases, we need richer structures like epistemic states in the sense of (Halpern 2005), most commonly being represented as probability distributions, possibility distributions (Dubois and Prade 2015) or Ordinal Conditional Functions (Spohn 1988; 2012). A knowledge base is *consistent* if and only if there is (a representation of) an epistemic state that accepts (all conditionals in) the knowledge base.

4 Plausible reasoning

In this section we recall inference based on plausibility relations as model for conditionals and framework of conditional logic. We furthermore recall how to generate an admissible plausibility relation to a conditional knowledge base by instantiating preferential models (Makinson 1994) with Ordinal Conditional Functions (OCF, (Spohn 1988; 2012)) which we inductively generate for the conditional knowledge base.

4.1 Preferential Inference

For nonmonotonic inference and the modeling of epistemic states, total preorders $\preccurlyeq$ on possible worlds expressing plausibility are of crucial importance. If $\omega_1 \preccurlyeq \omega_2$, ω_1 is deemed as at least as plausible as ω_2. Such a preorder can be lifted to the level of formulas by stating that $A \preccurlyeq B$ if for each model of B, there is a model of A that is at least as plausible. As usual, the relations $\prec$ and $\approx$ are derived from $\preccurlyeq$ by $A \prec B$ if and only if $A \preccurlyeq B$ and not $B \preccurlyeq A$, and $A \approx B$ if and only if both $A \preccurlyeq B$ and $B \preccurlyeq A$. Nonmonotonic inference can then be easily realized as a form of preferential entailment of high logical quality (Makinson 1994): $A|\!\sim_\preccurlyeq B$ if and only if $AB \prec A\overline{B}$, i.e., from A, B can be plausibly inferred if in the context of A, B is more plausible than $\overline{B}$. Hence total preorders provide convenient epistemic structures for plausible reasoning, and epistemic states Ψ can be represented by such a total preorder $\preccurlyeq_\Psi$. The belief set, i.e., the most plausible beliefs that an agent with epistemic state Ψ holds, is defined to be the set of all formulas which are satisfied by all most plausible worlds: $Bel(\Psi) = Th(\min(\preccurlyeq_\Psi))$, where $\min(\preccurlyeq_\Psi)$ is the set of all minimal worlds according to $\preccurlyeq$. Conditionals can then be integrated smoothly into this reasoning framework by defining $\Psi \models (B|A)$ if and only if $A|\!\sim_\preccurlyeq B$, i.e., conditionals can encode nonmonotonic inferences on the object level. We illustrate this with the following example, for more details, we refer to, e.g., (Kern-Isberner and Eichhorn 2014).

Example 1 *We illustrate this inference using the car example from (Thompson and Byrne 2002), so let G indicate that a car is out of gas (g), or not ($\overline{g}$), and S indicate that the car has stalled (s), or not ($\overline{s}$). Here the possible worlds are $\{gs, g\overline{s}, \overline{g}s, \overline{g}\,\overline{s}\}$. We define the epistemic state Ψ to be represented by the preorder*

$$gs \approx_\Psi \overline{g}\,\overline{s} \prec_\Psi g\overline{s} \approx_\Psi \overline{g}s.$$

Applying preferential inference we obtain that, for instance, $g|\!\sim_\preccurlyeq s$ because $gs \prec g\overline{s}$, thus $\Psi \models (s|g)$. Here, $\min(\preccurlyeq_\Psi) = \{gs, \overline{g}\,\overline{s}\}$, thus $Bel(\Psi) = Th(\{gs, \overline{g}\,\overline{s}\}) = Cn(g \Leftrightarrow s)$.

4.2 Ordinal Conditional Functions

Ordinal conditional functions (Spohn 2012) are specific implementations of such epistemic states that assign to each level of plausibility a degree of (im)plausibility. More formally, an Ordinal Conditional Function (OCF, (Spohn 1988; 2012)), also called a *ranking function*, is a function $\kappa : \Omega \rightarrow \mathbb{N}_0 \cup \{\infty\}$ that assigns to each world ω an implausibility rank $\kappa(\omega)$ such that the higher $\kappa(\omega)$, the less plausible ω is, and with the normalization constraint that there are worlds that are maximally plausible, that is, the pre-image $\kappa^{-1}(0)$ cannot be empty. The rank of a formula $A \in \mathfrak{L}$ is the minimal rank of all worlds that satisfy A, and the rank of a conditional is the rank of the verification of the conditional normalized by the rank of the premise, so we have $\kappa(A) = \min\{\kappa(\omega) | \omega \models A\}$ and $\kappa(B|A) = \kappa(AB) - \kappa(A)$.

A ranking function *accepts* a conditional (written $\kappa \models (B|A)$) if and only if its verification is more plausible than its falsification, and a formula B is κ-inferred from a formula A (written $A|\!\sim_\kappa B$) if and only if κ accepts the conditional $(B|A)$, if and only if $\kappa \models (B|A)$, if and only if $\kappa(AB) < \kappa(A\overline{B})$, in accordance with preferential inference as defined above. An OCF is *admissible with respect to a knowledge base* (written $\kappa \models \Delta$) if and only if it accepts all conditionals in Δ.

Example 2 *We continue Example 1 to illustrate OCF. A ranking function that induces $\preccurlyeq_\Psi$ is the OCF $\kappa(gs) = \kappa(\overline{g}\,\overline{s}) = 0$, $\kappa(g\overline{s}) = \kappa(\overline{g}s) = 1$. With κ we have $\kappa(gs) < \kappa(g\overline{s})$, and thus $\kappa \models (s|g)$ and also $g|\!\sim_\kappa s$.*

4.3 C-Representations

With admissible ranking functions we have a way to define an inference relation that is based on the knowledge represented in the knowledge base. We now go one step further and with the approach of c-representations recall an approach that for each consistent knowledge base Δ provides a schema for inductively setting up epistemic states in form of OCFs that are admissible with respect to Δ.

For this approach, we assign an integer impact κ_i^- to each conditional $(B_i|A_i) \in \Delta = \{(B_1|A_1), \ldots, (B_n|A_n)\}$, indicating how implausible it is to violate the conditional rule. A c-representation (Kern-Isberner 2001) is an OCF κ_Δ^c such

that the rank of each world is the sum of all impacts of conditionals in Δ which are falsified by this world, formally,

$$\kappa^c_\Delta(\omega) = \sum_{\omega \models A_i \overline{B}_i} \kappa_i^-, \quad (1)$$

where the impacts are chosen such that $\kappa^c_\Delta \models \Delta$, that is, $\kappa^c_\Delta(A_iB_i) < \kappa^c_\Delta(A_i\overline{B}_i)$. With the ranks of formulas and (1) for all $1 \leq i \leq n$ this constraint expands to

$$\min_{\omega \models A_iB_i} \Big\{ \underbrace{\sum_{\substack{\omega \models A_j \overline{B_j} \\ i=1}}^{n} \kappa_j^-}_{(2.a)} \Big\} \leq \min_{\omega \models A_i\overline{B}_i} \Big\{ \underbrace{\sum_{\substack{\omega \models A_j \overline{B_j} \\ i=1}}^{n} \kappa_j^-}_{(2.b)} \Big\}. \quad (2)$$

The left minimum ranges over the models of A_iB_i, so the conditional $(B_i|A_i)$ is not falsified by any considered world and thus κ_i^- is no element of any sum $(2.a)$. As opposed to this, the right minimum ranges over the models of $A_i\overline{B}_i$, so the conditional $(B_i|A_i)$ is falsified by every considered world and thus κ_i^- is an element of every sum $(2.b)$. With these deliberations, we can rewrite the inequalities to

$$\min_{\omega \models A_iB_i} \Big\{ \sum_{\substack{\omega \models A_j \overline{B_j} \\ i \neq j}} \kappa_j^- \Big\} \leq \kappa_i^- + \min_{\omega \models A_i\overline{B}_i} \Big\{ \sum_{\substack{\omega \models A_j \overline{B_j} \\ j \neq i}} \kappa_j^- \Big\}, \quad (3)$$

and therefore

$$\kappa_i^- \geq \min_{\omega \models A_iB_i} \Big\{ \sum_{\substack{\omega \models A_j \overline{B_j} \\ i \neq j}} \kappa_j^- \Big\} - \min_{\omega \models A_i\overline{B}_i} \Big\{ \sum_{\substack{\omega \models A_j \overline{B_j} \\ i \neq j}} \kappa_j^- \Big\} \quad (4)$$

for all $1 \leq i \leq n$. Via (1) every solution of this system of inequalities constitutes a ranking model of the knowledge base Δ. This means that Δ has ranking models and thus is consistent if and only if we can solve the system of inequalities (4) (confer (Kern-Isberner 2001)).

Example 3 *We further extend Example 1 and formalize the conditional statements of (Thompson and Byrne 2002, Experiment 2) as "If the car had been out of gas, then it (usually) would have stalled."* (δ_1) *and "If the car had not been out of gas, then it (usually) would not have stalled."* (δ_2) *as the knowledge base* $\Delta = \{\delta_1 : (s|g), \delta_2 : (\overline{s}|\overline{g})\}$*. Table 1 shows the verification / falsification behavior for this example. For* Δ*, the system of inequalities* (4) *yields*

$$\kappa_1^- > \min\{0\} - \min\{0\} \quad \rightarrow \quad \kappa_1^- > 0 \quad (5)$$
$$\kappa_2^- > \min\{0\} - \min\{0\} \quad \rightarrow \quad \kappa_2^- > 0. \quad (6)$$

The OCF κ_1 *originating from a minimal solution* $\kappa_1^- = 1$, $\kappa_2^- = 1$ *by means of* (1) *is given in Table 1. Here,* gs *is always more plausible as* $g\overline{s}$ *because the prior falsifies no conditionals and thus is assigned to a rank of* 0*, whilst the latter falsifies* δ_1 *which by* (5) *bestows the strictly positive impact* κ_1^-*. Therefore, all c-representations of this knowledge base license for the inference* $g|\sim^\Delta_{\kappa^c_\Delta} c$*, since we have* $\kappa^c_\Delta(gs) = 0$ *and* $\kappa^c_\Delta(g\,\overline{s}) = \kappa_1^- > 0$*. We also have* $\overline{s}|\sim^\Delta_{\kappa^c_\Delta} \overline{g}$ *because* $\kappa^c_\Delta(\overline{g}\,\overline{s}) = 0$ *and* $\kappa^c_\Delta(g\,\overline{s}) > 0$*.*

Table 1: Verification / falsification behavior and a c-representation for Example 3.

Worlds ω	$g\,s$	$g\,\overline{s}$	$\overline{g}\,s$	$\overline{g}\,\overline{s}$
verifies	δ_1	—	—	δ_2
falsifies	—	δ_1	δ_2	—
$\kappa_1(\omega)$	0	1	1	0

5 Inference patterns

Previous research discussed whether each of the inference rules MP, MT (both classically valid), and AC, DA (both classically invalid) is respectively applied by investigating if participants or any (cognitive) system draw a specific inference. Table 2 recalls the inference rules and their respective inferences. In this section, we go beyond that in two respects: first, we formalize what it means that it is *plausible* to draw conclusions according to these rules, and secondly, we focus on the combination of these inference rules in a specific experiment which reflects the global inference behavior typical for the respective experiment:

Definition 4 (Inference Pattern) *An* inference pattern ϱ *is a 4-tuple of inference rules that for each inference rule MP, MT, AC, and DA indicates whether the rule is used (positive rule, e.g., MP) or not used (negated rule, e.g.,* $\neg$*MP) in an inference scenario. The set of all 16 inference patterns is called* $\mathcal{R}$*.*

We illustrate these inference patterns and how they are associated with a set of inference rules used (and not used) by the participants with two experiments from the literature.

Example 5 (Suppression) *In the Suppression Task (Byrne 1989, Experiment 1 about additional arguments) the participants had to draw inferences with respect to the arguments "If Lisa has an essay to write, she will study late in the library." and "If the library stays open, she will study late in the library." Here, the majority of the participants, students without tuition in logic, did not apply MP* (38%) *or MT* (33%)*, but did apply AC* (63%) *and DA* (54%)*, which gives us the inference pattern* $\varrho_{B89} = (\neg MP, \neg MT, AC, DA)$*.*

Example 6 (Counterfactual) *In (Thompson and Byrne 2002, Experiment 2, reasoning about nonnecessary/causal) the participants had to reason about counterfactual statements like "If the car had been out of gas, then it would have stalled". Here, the majority of the participants, who were introductory students of psychology, applied MP* (78%) *and MT* (85%)*, and did not apply AC* (41%)*. They also tended to apply DA* (50%)*, thus we determine this study can be covered by the pattern* $\varrho_{TB02} = (MP, MT, \neg AC, DA)$*. Note that, as the participants are on the fence with DA, another option is to formalize this as the pattern* $\varrho_{TB02'} = (MP, MT, \neg AC, \neg DA)$*.*

To draw plausible inferences with respect to an inference rule, a plausibility preorder $\preccurlyeq$ has to be defined on the set of worlds see Section 4. For instance, we have MP if any only if for a statement "If A then B" the inference $A|\sim B$ is drawn. This is the case if and only if the worlds are ordered such

Table 2: Overview of the inferences drawn or not drawn from "From A it (usually) follows that B" with respect to appliance of the inference rules.

Rule	Inference	Rule	Inference
MP	$A \mid\!\sim B$	¬MP	$A \mid\!\not\sim B$
MT	$\overline{B} \mid\!\sim \overline{A}$	¬MT	$\overline{B} \mid\!\not\sim \overline{A}$
AC	$B \mid\!\sim A$	¬AC	$B \mid\!\not\sim A$
DA	$\overline{A} \mid\!\sim \overline{B}$	¬DA	$\overline{A} \mid\!\not\sim \overline{B}$

Table 3: Constraints on the plausibility relation on worlds in order to satisfy inference rules.

Rule	Plausibility constraint	Rule	Plausibility constraint
MP	$AB \prec A\overline{B}$	¬MP	$A\overline{B} \preccurlyeq AB$
MT	$\overline{A}\,\overline{B} \prec A\overline{B}$	¬MT	$A\overline{B} \preccurlyeq \overline{A}\,\overline{B}$
AC	$AB \prec \overline{A}B$	¬AC	$\overline{A}B \preccurlyeq AB$
DA	$\overline{A}\,\overline{B} \prec \overline{A}B$	¬DA	$\overline{A}B \preccurlyeq \overline{A}\,\overline{B}$

that for each world violating the statement (each $\omega' \models A\overline{B}$) there is a world that verifies the statement ($\omega \models AB$) which is more plausible than ω' ($\omega \prec \omega'$), that is, if and only if $AB \prec A\overline{B}$. Table 3 gives all of the plausibility constraints which are equivalent to using the inference rules.

To satisfy an inference pattern, the plausibility relation has to satisfy each of the constraints given in Table 3. So each reasoning pattern $\varrho \in \mathcal{R}$ imposes a set of constraints on the plausibility relation, which in the following is called $\mathcal{C}(\varrho)$; $\mathcal{C}(\varrho)$ is *satisfiable* if and only if there is a plausibility relation $\prec$ and hence an epistemic state that satisfies all constraints in $\mathcal{C}(\varrho)$. For instance, to satisfy the pattern ϱ_{TB02} from Example 6, the worlds have to be ordered such that all four constraints given in Table 4 are satisfied. If for a given pattern ϱ, there is a plausibility relation $\preccurlyeq$ that satisfies $\mathcal{C}(\varrho)$, that is, there is a preorder on the worlds which is in accordance with plausible reasoning, ϱ can be deemed to be rational. Therefore we define a plausibility and constraint-based notion of rationality for inference patterns as follows:

- An inference pattern $\varrho \in \mathcal{R}$ is *rational* if and only if there is a plausibility relation $\prec \subseteq \Omega \times \Omega$ that satisfies $\mathcal{C}(\varrho)$.
- Otherwise, the inference pattern is *irrational*.

Inspecting all $\varrho \in \mathcal{R}$ we obtain that only two patterns, namely $(\text{MP}, \neg\text{MT}, \neg\text{AC}, \text{DA})$ and $(\neg\text{MP}, \text{MT}, \text{AC}, \neg\text{DA})$, are irrational: For the first pattern, the constraints impose the unrealizable ordering $\overline{A}\,\overline{B} \prec \overline{A}B \preccurlyeq AB \prec A\overline{B} \preccurlyeq \overline{A}\,\overline{B}$, for the second, the constraints impose the unrealizable ordering $\overline{A}\,\overline{B} \prec A\overline{B} \preccurlyeq AB \prec \overline{A}B \preccurlyeq \overline{A}\,\overline{B}$.

6 Explaining Human Inferences

The previous section showed that, apart from the two irrational inference patterns $(\text{MP}, \neg\text{MT}, \neg\text{AC}, \text{DA})$ and $(\neg\text{MP}, \text{MT}, \text{AC}, \neg\text{DA})$, there are plausibility relations on the possible worlds for all other inference patterns in $\mathcal{R}$ and

Table 4: Constraints for the inference pattern $\varrho_{TB02} = (\text{MP}, \text{MT}, \neg\text{AC}, \text{DA})$ (written $\mathcal{C}(\varrho_{TB02})$).

$\{AB \prec A\overline{B}, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B \preccurlyeq AB, \overline{A}\,\overline{B} \prec \overline{A}B\}$
$\equiv \quad \overline{A}\,\overline{B} \prec \overline{A}B \preccurlyeq AB \prec A\overline{B}$

hence these inference pattern can be deemed rational according to the standards of plausible reasoning. In this section we show how these plausibility relations induced by the constraints imposed by inference patterns can be used to algorithmically set up conditional knowledge bases as formal explanations for the inferences and thus to generate hypotheses about the background knowledge used by the participants. Moreover, from the plausibility relations, we can also extract the most plausible beliefs which might reveal implicit assumptions used by the participants. First, we have to extend our framework of conditionals and c-representations a bit.

We defined an inference pattern to consist of inferences and non-inferences, the latter being equivalent to the non-acceptance of a conditional by a ranking functions, or a non-strict constraint on the plausibility preorder. For instance, ¬MT implies the constraint $A\overline{B} \preccurlyeq \overline{A}\,\overline{B}$ and thus the non-acceptance $\kappa \not\models (\overline{A}|\overline{B})$. This is *not* equivalent to the acceptance of the negation of the conditional ($\kappa \models (A|\overline{B})$), but weaker, because it also allows for indifference between both cases AB and $A\overline{B}$. As a consequence neither the conditional nor its negation may be accepted. We capture this effect with so-called *weak conditionals*:

Definition 7 (Weak conditional) *Let $A, B \in \mathfrak{L}$ be formulas. We define $(\!|B|A|\!)$ to be a* weak conditional *with the same three-valued evaluation as given in Section 3, but with the semantics that a ranking function (or a plausibility preorder) accepts $(\!|B|A|\!)$ if and only if it does not accept its negation; more precisely, for a ranking function κ,*

$$\kappa \models (\!|B|A|\!) \;\; \textit{iff} \;\; \kappa \not\models (\overline{B}|A) \;\; \textit{iff} \;\; \kappa(AB) \leq \kappa(A\overline{B}). \quad (7)$$

The language of all weak conditionals is denoted by $(\!|\mathfrak{L}|\mathfrak{L}|\!)$.

In the following, conditionals $(B|A) \in (\mathfrak{L}|\mathfrak{L})$ are referred to as *strong* conditionals, if necessary. Our knowledge bases will consist of both strong and weak conditionals in the rest of this paper.

Let Δ be a knowledge base containing the weak conditional $(\!|B_o|A_o|\!)$. To apply c-representations to Δ, the impact κ_o^- has to be chosen such that κ_Δ^c satisfies (7), that is, $\kappa_\Delta^c(A_oB_o) \leq \kappa_\Delta^c(A_o\overline{B}_o)$. Applying the definition of ranks of formulas and the definition of κ_Δ^c in (1), this expands to

$$\min_{\omega \models A_oB_o} \Big\{ \sum_{\substack{\omega \models A_i\overline{B_i} \\ i=1}}^{n} \kappa_i^- \Big\} \leq \min_{\omega \models A_o\overline{B}_o} \Big\{ \sum_{\substack{\omega \models A_i\overline{B_i} \\ i=1}}^{n} \kappa_i^- \Big\}. \quad (8)$$

With the same steps that lead from (2) to (4) we obtain

$$\kappa_o^- \geq \min_{\omega \models A_oB_o} \Big\{ \sum_{\substack{\omega \models A_i\overline{B_i} \\ i \neq o}} \kappa_i^- \Big\} - \min_{\omega \models A_o\overline{B}_o} \Big\{ \sum_{\substack{\omega \models A_i\overline{B_i} \\ i \neq o}} \kappa_i^- \Big\} \quad (9)$$

Table 5: Strong and weak conditionals for the inference rules.

Rule	Conditional	Rule	Conditional
MP	$(B\|A)$	¬MP	$(\!(\overline{B}\|A)\!)$
MT	$(\overline{A}\|\overline{B})$	¬MT	$(\!(A\|\overline{B})\!)$
AC	$(A\|B)$	¬AC	$(\!(\overline{A}\|B)\!)$
DA	$(\overline{B}\|\overline{A})$	¬DA	$(\!(B\|\overline{A})\!)$

which equals (4) save for the strictness of the inequality. So to apply c-representations to a conditional knowledge base containing weak and strong conditionals, we define a system of inequalities (4') according to the schema (4) where the inequality for the impact κ_i^- is strict if δ_i is a strong conditional and not strict, otherwise.

With weak conditionals and (4') we can now propose an algorithm to generate a most concise formal explanation for the plausibility relation imposed by the constraints of an inference pattern. This algorithm takes initially a knowledge base with one conditional for each rule in the inference pattern. Then the constraint system (4') is set up from this knowledge base and analyzed with respect to redundancies. (Weak) Conditionals the appertaining inequality of which is implied by the other inequations are removed from the knowledge base. This way, the only conditionals remaining in the knowledge base are those whose effects on the epistemic state represented by a corresponding c-representation are most relevant.

Algorithm: Explanation Generator
Input: Inference pattern $\varrho \in \mathcal{R}$
Output: Knowledge base $\Delta \subseteq (\mathfrak{L}|\mathfrak{L}) \cup (\!(\mathfrak{L}|\mathfrak{L})\!)$

1. Set up Δ^* with a conditional for each rule in pattern ϱ according to Table 5.
2. Set up the system of inequalities (4') for Δ^* and simplify: For each inequality that is implied by the other inequalities, remove the line from the system of inequalities and the respective conditional from Δ^* to obtain the (wrt. set inclusion) *minimal explaining knowledge base* Δ.
3. Return the knowledge base Δ.

We illustrate this algorithm by finding formal explanations and thus hypotheses for background knowledge:

Example 8 *The inferences in the Suppression Task (Example 5) can be captured by the inference pattern* $\varrho_{B89} = (\neg MP, \neg MT, AC, DA)$. *Our modeling revolves around the relationship between the antecedent and consequent of the query, not the background information. Thus the algorithm deliberately works with conditionals concerning relationships between literals of L ("Lisa (not) being in the library") and E ("Lisa (not) having an essay to write") and not O ("the library (not) being open"), even if information about this was present in the cover story. For the pattern* ϱ_{B89}, *the algorithm sets up the knowledge base* $\Delta^*_{B89} = \{\delta_1 : (\!(\overline{l}|e)\!), \delta_2 : (\!(e|\overline{l})\!), \delta_3 : (e|l), \delta_4 : (\overline{l}|\overline{e})\}$ *from the conditionals* $(\!(\overline{l}|e)\!)$ *because* $\neg MP \in \varrho_{B89}$, $(\!(e|\overline{l})\!)$ *because* $\neg MT \in \varrho_{B89}$, *and* $(e|l)$ *and* $(\overline{l}|\overline{e})$ *because AC and DA are in* ϱ_{B89}. *The verification / falsification behavior of worlds with respect to* Δ^*_{B89} *is shown in Table 6. As a next step, we set up and simplify the system of inequalities (4') for* Δ^*_{B89}*:*

$$\kappa_1^- \geq 0, \quad \kappa_2^- \geq 0, \quad \kappa_3^- > \kappa_1^- - \kappa_4^-, \quad \kappa_4^- > \kappa_2^- - \kappa_3^-$$

Here the inequalities for κ_1^- *and* κ_2^- *do not yield additional information to the definition of* κ_i^- *being an element of* $\mathbb{N}_0$, *thus these two lines (and the respective conditionals) can be removed, which leaves us with the inequality* $\kappa_3^- + \kappa_4^- > \max\{\kappa_1^-, \kappa_2^-\}$. *For or a minimal (i.e., most conservative) solution of this inequality, it is sufficient if one of the conditionals has a positive impact. So the background knowledge that can formally explain this pattern is one of the knowledge bases* $\Delta_{B89} = \{(e|l)\}$ *and* $\Delta'_{B89} = \{(\overline{l}|\overline{e})\}$. *These two conditionals are falsified by the same world, so the c-representations for both knowledge bases are equivalent with respect to the induced ranking functions and plausibility relations; Table 6 gives a minimal instantiation* κ_2. *Nonetheless both* Δ_{B89} *and* Δ'_{B89} *are formal explanations for the inference pattern* ϱ_{B89}. *This gives us the hypothesis that the background knowledge used by the majority of the participants in this task was either the conditional*

δ_3 *"If Lisa is in the library, then she (usually) has an essay to write",*

or the conditional

δ_4 *"If Lisa has no essay to write, then she (usually) is not in the library".*

Regarding the explanations Δ_{B89} and Δ'_{B89}, the inference pattern found in Example 5 appears to be rational: The participants might have understood the given conditional information in its inverse form, and hence applied AC and DA which in fact, amount to MP and MT for the inverse conditional. Also the most plausible beliefs $Bel(\kappa^c_{\Delta_{B89}}) = Cn(e \vee \overline{l}) = Cn(l \Rightarrow e)$ reveal the implicit assumption that l should imply e (or $\overline{e}$ should imply $\overline{l}$).

Table 6: Verification / falsification behavior and one c-representation for the knowledge base Δ^*_{B89} of the Suppression Task.

ω	$e\,l$	$e\,\overline{l}$	$\overline{e}\,l$	$\overline{e}\,\overline{l}$
verifies	δ_3	δ_1,δ_2	—	δ_4
falsifies	δ_1	—	δ_3,δ_4	δ_2
$\kappa_2(\omega)$	0	0	1	0

Example 9 *In Example 6, the participants reasoned according to the pattern* $\varrho_{TB02} = (MP, MT, \neg AC, DA)$ *which yields the constraints in Table 4. For this pattern the algorithm sets up the knowledge base* $\Delta^*_{TB02} = \{\delta_1 : (s|g), \delta_2 : (\overline{g}|\overline{s}), \delta_3 : (\!(\overline{g}|s)\!), \delta_4 : (\overline{s}|\overline{g})\}$, *which, according to (4'), yields the following system of inequalities*

$$\kappa_1^- > \kappa_3^- - \kappa_2^- \qquad \kappa_2^- > 0 - \kappa_1^-$$
$$\kappa_3^- \geq \kappa_4^- - 0 \qquad \kappa_4^- > 0 - 0.$$

Table 7: Verification / falsification behavior for the algorithm Explanation Generator in Example 9 (alternative modelling) and resulting OCF κ_3.

ω	$g\,s$	$g\,\overline{s}$	$\overline{g}\,s$	$\overline{g}\,\overline{s}$
verifies	δ_1	—	δ_3,δ_4	δ_2
falsifies	δ_3	δ_1,δ_2	—	δ_4
$\kappa_3(\omega)$	0	1	0	0

Here the impact of δ_2 is covered by the other conditionals, so the system can be simplified to $\kappa_1^- > \kappa_3^-$, $\kappa_3^- \geq \kappa_4^-$, and $\kappa_4^- > 0$, which gives us the minimal explaining knowledge base $\Delta_{TB02} = \{\delta_1 : (s|g), \delta_3 : (\overline{g}|s), \delta_4 : (\overline{s}|\overline{g})\}$. Using the techniques defined in Section 4 and demonstrated in Example 3 it can be seen that every c-representation for Δ_{TB02} satisfies the constraints of $\mathcal{C}(\varrho_{TB02})$. From Δ_{TB02} we read the hypothesis for the background knowledge to be:

δ_1 *"If the car had been out of gas, then (usually) it would have stalled."*

δ_3 *"If the car had stalled, then it possibly would not have been out of gas"*[1]

δ_4 *"If the car had not been out of gas, then (usually) it would not have stalled."*

We also apply this method to the alternative modeling of the study in Example 6, that is, the inference pattern $\varrho_{TB02'} = (MP, MT, \neg AC, \neg DA)$. Using the algorithm Explanation Generator gives us the initial knowledge base $\Delta^*_{TB02'} = \{\delta_1 : (s|g), \delta_2 : (\overline{g}|\overline{s}), \delta_3 : (\overline{g}|s), \delta_4 : (s|\overline{g})\}$. Table 7 (upper rows) shows the verification / falsification behavior for the worlds in this example. Here (4') is instantiated to

$$\kappa_1^- > \kappa_3^- - \kappa_2^-, \quad \kappa_2^- > \kappa_4^- - \kappa_1^-, \quad \kappa_3^- \geq 0, \quad \kappa_4^- \geq 0$$

which the algorithm simplifies to $\kappa_1^- + \kappa_2^- > 0$. Only one of these impacts has to be positive to satisfy the inequality, which means that the others (together with the respective conditionals) can be removed. This results in the hypotheses $\Delta_{TB02'} = \{(s|g)\}$ and $\Delta'_{TB02'} = \{(\overline{g}|\overline{s})\}$ as explanations for the inferences, and a minimal c-representation κ_3 given in Table 7 (bottom row). Here the belief set is $Bel(\Delta_{TB02'}) = Cn(gs, \overline{g}s, \overline{g}\,\overline{s}) = Cn(g \Rightarrow s)$. This result shows that the pattern indicating the usage of only classically valid inference rules suitably characterizes the classical logic reasoner that takes a conditional as a material implication.

The inferences drawn with respect to the pattern ϱ_{TB02} in Example 9 can be considered rational in the light of the explanation Δ_{TB02}: The participants might have understood the counterfactual "If the car had been out of gas, then it would have stalled." as the two conditionals δ_1 and δ_4, applying MP, MT and DA, and additionally assumed there might be additional reasons for a car to stall which are not mentioned in the task, thus preventing them from applying AC; this assumption is encoded as the weak conditional δ_3. The most plausible beliefs $Bel(\Delta_{TB02}) = Cn(\overline{g}\,\overline{s})$ reveal the implicit assumption that neither did the car stall, nor has it been out of gas in accordance of this being a task of counterfactual reasoning, that is, reasoning about things that are (or were) not true.

Note that the weak conditional δ_3 is necessary to explain this inference pattern: In Example 3, we set up a c-representation from $\{(s|g), (\overline{s}|\overline{g})\}$, that is, Δ_{TB02} without δ_3, which results in the inference pattern $(\text{MP}, \text{MT}, \text{AC}, \text{DA})$.

[1] Alternatively: "It is not true that if the car had stalled, it would (usually) have been out of gas."

7 Summary and Conclusion

When dealing with conditional statements "If A then B", human reasoners can apply some of the four inference rules MP, MT, AC, and DA, each of these rules representing a non-monotonic inference (e.g., $A|\sim B$ for MP). We proposed an approach to join the four rules into tuples called *inference patterns* which allows for classifying psychological findings, for instance, $(\text{MP}, \text{MT}, \neg\text{AC}, \text{DA})$ for (Thompson and Byrne 2002, Experiment 2), or the inference type of the individual participant. Non-monotonic inferences can be implemented by plausibility relations on possible worlds, yielding a notion of preferential entailment of a good logical quality. In this paper, *Ordinal Conditional Functions* (OCF, (Spohn 2012)) were used for defining such a plausibility on the worlds, and c-representations as an inductive approach to generate OCFs from conditional knowledge bases.

To draw inferences with respect to a pattern and plausible reasoning, the plausibility relation has to be in such a way that each of the represented non-monotonic inferences can be drawn. Thus, an inference pattern imposes constraints on the plausibility relation over the possible worlds. We have shown that for some patterns these constraints are not satisfiable, and, hence, can be considered as irrational with respect to plausibility reasoning. Examining the plausibility relations, we were able to identify implicit assumptions as the most plausible beliefs: In the counterfactual case of (Thompson and Byrne 2002), the induced belief set is $Cn(\overline{A}\,\overline{B})$; for the classically valid pattern $(\text{MP}, \text{MT}, \neg\text{AC}, \neg\text{DA})$, on the other hand, the most plausible worlds are the models of the material implication, that is, the belief set is $Cn(A \Rightarrow B)$, which supports the interpretation that the reasoner indeed has understood the conditional as an implication.

By introducing weak conditionals, we captured each rational pattern as a conditional knowledge base. Using c-representations we reduced these knowledge bases to minimal knowledge bases, thus generating a compact formal explanation for the inferences in the pattern. From such knowledge bases we derived hypotheses about the background knowledge of the participants in the experiments. This whole process is captured in an algorithm which systematically constitutes such hypotheses for conditional inference tasks in cognitive science.

Summing up the paper, we applied a logic based on conditionals and plausible reasoning to evaluate inferences found in experiments according to logical standards. We emphasized that it is necessary to consider all (non-)applied inference rules as a pattern which then can be represented

and solved as constraints on plausibility relations. Patterns for which these constraints are jointly solvable, that is, a plausibility-based representation of an epistemic state that incorporates all inference rules exists, are deemed to be *rational* in our approach. In this, we laid a cognitive and formal foundation for assessing "rationality of a reasoning system" beyond mere philosophical reflection into the domain of (mathematical / algorithmic) constraint solving and explained why humans deviating in the applied inference schemata (for instance MP and AC) from classical logic (allowing only MP and MT as correct) are not necessarily wrong, as long as the reasoning schema is inherently consistent. This way, our approach offers new insights into the phenomenon of rationality and novel techniques to evaluate what may be called rational, and what might rather be called irrational. It provides the ground for a possible paradigm shift, namely that cognitive science researcher start to evaluate complete inference systems with respect to their internal consistency according to conditional logic rather than inferences alone. Finally, the formal approach on psychological studies and their results provides a common ground for comparing rationality of both AI and human reasoning on relevant paradigmatic reasoning examples.

Overall the method provides a new approach that evaluates for each system whether the application of reasoning schemata contradict each other. The approach can be applied to single reasoners and their internal consistency by the same method. This paper also provides methods to systematically reveal conditional knowledge bases representing background knowledge, and implicit assumptions on which the observed inference patterns can be based on and thus explained. As part of our current research we plan to experimentally evaluate the approach, in that we determine the inference pattern of the individual participants and then ask whether they accept the hypothesis calculated by the Explanation Generator as (part of) their background knowledge.

Acknowledgements

This work is supported by DFG-Grants KI1413/5-1 to G. Kern-Isberner and RA1934 2/1 as part of the priority program "New Frameworks of Rationality" (SPP 1516), and a Heisenberg DFG fellowship RA1934 3/1, 4/1 to M. Ragni. C. Eichhorn is supported by Grant KI1413/5-1.

References

Baratgin, J.; Over, D. E.; and Politzer, G. 2013. Uncertainty and the de Finetti tables. *Thinking and Reasoning* 19(3–4):308–328.

Bonnefon, J.-F., and Hilton, D. J. 2004. Consequential conditionals: invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(1):28.

Byrne, R. M., and Egan, S. M. 2004. Counterfactual and prefactual conditionals. *Canadian Journal of Experimental Psychology* 58(2):113–120.

Byrne, R. M., and Tasso, A. 1999. Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory and Cognition* 27(4):726–740.

Byrne, R. M. 1989. Suppressing valid inferences with conditionals. *Cognition* 31:61–83.

Castañeda, L. E. G., and Knauff, M. 2016. Defeasible reasoning with legal conditionals. *Memory & cognition* 44(3):499–517.

De Neys, W.; Schaeken, W.; and D'Ydewalle, G. 2003. Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition* 31(4):581–595.

DeFinetti, B. 1974. *Theory of Probability: A Critical Introductory Treatment (Translated by A. Machi and A. Smith)*, volume 1 & 2. UK: Wiley.

Dubois, D., and Prade, H. 2015. Possibility theory and its applications: Where do we stand? In Kacprzyk, J., and Pedrycz, W., eds., *Springer Handbook of Computational Intelligence*. Berlin, DE: Springer. 31–60.

Egan, S. M., and Byrne, R. M. 2012. Inferences from counterfactual threats and promises. *Experimental psychology*.

Egan, S. M.; García-Madruga, J. A.; and Byrne, R. M. 2009. Indicative and counterfactual 'only if' conditionals. *Acta psychologica* 132(3):240–249.

Frosch, C. A., and Byrne, R. M. 2012. Causal conditionals and counterfactuals. *Acta psychologica* 141(1):54–66.

Griggs, R. A., and Cox, J. R. 1982. The elusive thematic-materials effect in wason's selection task. *British Journal of Psychology* 73(3):407–420.

Halpern, J. Y. 2005. *Reasoning About Uncertainty*. Cambridge, MA, USA: MIT Press.

Kern-Isberner, G., and Eichhorn, C. 2014. Structural Inference from Conditional Knowledge Bases. In Unterhuber, M., and Schurz, G., eds., *Logic and Probability: Reasoning in Uncertain Environments*, number 102 (4) in Studia Logica. Dordrecht, NL: Springer Science+Business Media. 751–769.

Kern-Isberner, G. 2001. *Conditionals in Nonmonotonic Reasoning and Belief Revision: Considering Conditionals as Agents*. Lecture Notes in Computer Science. Springer Science+Business Media. Springer Berlin Heidelberg.

Makinson, D. 1994. General Patterns in Nonmonotonic Reasoning, vol. 3. In Gabbay, D. M.; Hogger, C. J.; and Robinson, J. A., eds., *Handbook of Logic in Artificial Intelligence and Logic Programming*. Oxford Uni. Press. 35–110.

Politzer, G. 2005. Uncertainty and the suppression of inferences. *Thinking & Reasoning* 11(1):5–33.

Quelhas, A. C., and Byrne, R. 2003. Reasoning with deontic and counterfactual conditionals. *Thinking and Reasoning* 9(1):43–65.

Ragni, M.; Eichhorn, C.; and Kern-Isberner, G. 2016. Simulating human inferences in the light of new information: A formal analysis. In Kambhampati, S., ed., *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 2604–10.

Spohn, W. 1988. Ordinal Conditional Functions: A Dynamic Theory of Epistemic States. In *Causation in Decision, Belief Change and Statistics: Proceedings of the Irvine Conference on Probability and Causation*, volume 42 of *The Western Ontario Series in Philosophy of Science*, 105–134. Dordrecht, NL: Springer Science+Business Media.

Spohn, W. 2012. *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford, UK: Oxford University Press.

Thompson, V. A., and Byrne, R. M. 2002. Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(6):1154–1170.

Wason, P. C. 1968. Reasoning about a rule. *Quarterly Journal of Experimental Psychology* 20(3):273–81.