On the Cognitive Adequacy of Non-monotonic Logics

Sara Todorovikj[1,2] , **Gabriele Kern-Isberner**[3] , **Marco Ragni**[1,2]
[1]Cognitive Computation Lab, Technical Faculty, University of Freiburg
[2]Danish Institute for Advanced Study, University of Southern Denmark
[3]Technical University Dortmund
{satod, ragni}@sdu.dk, gabriele.kern-isberner@cs.uni-dortmund.de

Abstract

Humans have the ability to reason conditionally despite the existence of disablers. They have the capability to consider content and background knowledge and they are prototypical non-monotonic reasoners. So far most research has focused on explaining an "average" reasoner and neglected the individual reasoning process. Towards identifying the specifics of human reasoning, we investigate the inference mechanism for conditional reasoning considering experimental data presented by a previous psychological study. The experimental material included a range of different problems and contents with varying amounts of disablers and alternatives. We consider individual inference patterns and explain them by a ranking on worlds and ordinal conditional functions. We investigate: (i) Do effects found on aggregate level still hold on the individual level, and if yes - to which extent? (ii) How can possible disablers and alternatives change the inference pattern? (iii) How do individuals differ among each other and are there any common patterns? With this analysis we show how non-monotonic logic provides a suitable tool to express and explain the specifics of human reasoning formally in a more coherent way than classical logic.

1 Introduction

You are given the following information (Singmann, Klauer, and Beller 2016):

If a balloon is pricked with a needle, then it will pop.
A balloon is pricked with a needle.

Then, you are asked to answer the following question with an *endorsement* in the form of a *probability value* between 0% and 100%:

How likely is it that it will pop?

Given the information you are provided with, and no reason to believe otherwise, your answer would most likely tend towards 100%. However, in this world of balloons and needles, consider the following information:

The balloon is without air, i.e., empty.

If you mentally consider such situations where the balloon would *not* pop, then your endorsement will most likely be lower than 100%. States like this are called *disablers*. On the other hand, there can be additional cases, e.g.:

The balloon is pricked with a pen.

that are called *alternatives*.

Depending on the different scenarios, in the form of disablers and alternatives, that an individual knows about and can think of, their endorsements can vary to a great extent. E.g., in the balloon scenario, after considering the information that the balloon might be without air, your answer might be 95% instead of 100%. Another person, due to their own personal background, might consider that information as more influential, so they would answer with e.g., 80%. In human reasoning literature many have focused on aggregating over an experiment's participants and just explaining the most frequently given answers. However, examples like this point to the need for an analysis shift to the individual level.

Table 1: Conditional Inference Forms

Premise	MP	AC	DA	MT
Major	X→Y	X→Y	X→Y	X→Y
Minor	X	Y	¬X	¬Y
Conclusion	Y	X	¬Y	¬X

Conditionals are statements usually of the form "If X then Y", or equivalently "Y, if X" (also written as X→Y, where X is the *antecedent* and Y the *consequent*). Conditionals are relevant in everyday life and science to describe causal, counterfactual, and other forms of relations between two propositions X and Y. By combining a conditional (also called a *major premise*) with a current state of a proposition (also called a *minor premise*), a *conclusion* can be inferred about the state of the other proposition. There are four major inference forms: *modus ponens* (MP), *modus tollens* (MT), *affirming the consequent* (AC) and *denying the antecedent* (DA), as shown in Table 1. Humans systematically deviate from interpreting conditionals as material implication (Ragni, Kola, and Johnson-Laird 2018). Despite more than 50 years of research there is still no cognitive theory that can fully explain human conditional reasoning processes and effects recognized through experimental data (Ragni, Dames, and Johnson-Laird 2019). Learning more about how individuals interpret different conditional reasoning tasks is crucial in order to take a step forward towards understanding human reasoning.

As motivated earlier, in this paper, we turn to the individual participant. We are interested in the inference process of the individual, we want to investigate which conclusion depending on the inference types this individual endorses and which not and how this individual differs in applying these inference mechanisms.

Eichhorn, Kern-Isberner, and Ragni (2018) propose an inference analysis approach using *inference patterns* based upon conditional logic. Contrary to previous research that largely focuses on the inference forms individually, they joined all inferences into one tuple. Here, we propose an enhancement that allows for inference patterns to be applied on *probabilistic* experimental data, by also taking into account the relationships between the inference form *endorsements*. Similarly, here the rationality of an inference pattern is assessed based on a plausibility semantics derived from preferential models (Makinson 1994) respectively Ordinal Conditional Functions (OCF, (Spohn 1988)), allowing for a deviation from following logical inference rules. Following this idea, we propose total preorders over possible worlds as *preferential mental models*, serving as cognitive models for reasoning of humans when they are presented with a conditional reasoning task.

In this way, we combine basic approaches from nonmonotonic reasoning and cognitive science on a deep methodological level to set up a formal framework of human reasoning that goes beyond classical logic, but does not need quantifications via, e.g., probabilities in the first place. These preferential mental models can be applied on the level of individuals, as well as on an aggregated level, to reveal basic structures of reasoning.

A *mental model* consists of the true states of the propositions in a premise. Given a conditional premise "If A then B", its mental model representation would consider the states of the propositions A and B. One of the most prominent reasoning theories that uses mental models is the Mental Model Theory (Johnson-Laird and Byrne 1991; Johnson-Laird and Byrne 2002). It assumes that when presented with a conditional, individuals start with an initial model where both propositions are true:

A B
...

Once the initial mental model is created, it triggers the recollection of relevant facts and background knowledge related to the conditional premise (Johnson-Laird and Byrne 1991). With that the initial model is either confirmed as correct or it stimulates the individual to engage in search for counterexamples, which would lead to the so-called *fleshed-out* representation, consisting of all states for which the conditional *holds*:

A	B
¬A	¬B
¬A	B

However, an interesting question is – given that there are limited cognitive capacities, which models are constructed and which are neglected? Is it possible to reverse-engineer the underlying rank of models and identify the preferred mental models? Which influence do alternatives or disablers have? Is it possible to formally found the Mental Model Theory (Johnson-Laird and Byrne 2002)? This will be investigated in the paper.

The paper is structured as follows. In the next section we will provide the empirical bases, an experimental study conducted by Singmann et al. (2016). Then we introduce some formal preliminaries and introduce plausible reasoning. In Section 5 we introduce the formal foundation for inference patterns and in Section 6 how they can be extended towards endorsements. In Section 7 we explain the empirical results with the formal framework we have developed. Section 8 discusses and concludes the article.

2 Experimental Data

Singmann et al. (2016) present four experiments in which they studied endorsement rates of the respective conclusions for the four inference forms. In three of them they use contents with a varying amount of disablers and alternatives. The fourth experiment manipulates the speaker expertise and differs from the others, hence, we do not consider it here.

The experimental data by Singmann et al. (2016)[1] considered here is from Experiments 1, 3a and 3b. In all three experiments, participants gave endorsements for the four inference forms. The contents are the same in all three experiments and they vary in the amounts of disablers and alternatives associated with them, quantified with 'Few' and 'Many', as shown in Table 2. Moreover, in Experiments 3a and 3b, participants were divided in three groups. In two of them they are given additional information in the form of disablers and alternatives, whereas in the last group participants received only the conditional task. Participants were asked to endorse the conclusion as a probability in the range 0 - 100%.

Each content is presented as a reduced inference (no major premise), e.g., for MP:

A balloon is pricked with a needle.
———
How likely is it that it will pop?

and additionally as a full conditional inference.

In the original study, Singmann et al. (2016) aggregate the participants from all three experiments, which is the approach that we also follow here. The number of participants in Exp. 1 is N = 31, in Exp. 3a is N = 77 and Exp. 3b is N = 91, making the total N = 199.

Table 3 presents the average endorsement values among participants for each inference form for all contents in both conditional presentation forms (reduced and full inference).

3 Formal Preliminaries

Building up on Eichhorn et al. (2018), we base our formal modeling approach 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

[1] The data can be found at https://osf.io/zcdfq

Table 2: Contents used in Singmann et al. (2016) experiments. *Note:* This is a translation of the contents in English as provided by the authors. The experiment has been conducted in German.

Keyword	Content	Disablers	Alternatives
Predator	If a predator is hungry, then it will search for prey.	Few	Few
Balloon	If a balloon is pricked with a needle, then it will pop.	Few	Many
Girl	If a girl has sexual intercourse, then she will be pregnant.	Many	Few
Coke	If a person drinks a lot of coke, then the person will gain weight.	Many	Many

Table 3: Average inference form endorsements for each task in each conditional presentation form from Singmann et al.'s (2016) experimental data. ('Red.' - Reduced Inference, 'Full' - Full Inference, 'Dis' - Disablers, 'Alt' - Alternatives, 'F' - Few, 'M' - Many)

Form	Task	Dis/Alt	MP	AC	DA	MT
Red.	Predator	F/F	91	84	74	81
	Balloon	F/M	89	64	74	81
	Girl	M/F	32	88	85	44
	Coke	M/M	64	52	57	60
Full	Predator	F/F	92	87	80	85
	Balloon	F/M	93	77	76	86
	Girl	M/F	62	87	83	62
	Coke	M/M	78	64	63	73

not ($\neg$), as usual. For simplicity, the symbol $\wedge$ might be omitted and the conjunction would be written by juxtaposition. Additionally, the negation ($\neg A$) would be abbreviated 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, $[\![A]\!]_\omega = \mathit{true}$ if and only if $\omega \models A$ if and only if $\omega \in \mathit{Mod}(A)$, that is, ω is an element of the classical models $\mathit{Mod}(A)$ of A. The set of classical consequences of a set of formulas $\mathcal{A} \subseteq \mathfrak{L}$ is $\mathit{Cn}(\mathcal{A}) = \{B | \mathcal{A} \models B\}$. 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 $\mathit{Th}(\mathcal{W}) = \{A \in \mathfrak{L} \mid \omega \models A \text{ for all } \omega \in \mathcal{W}\}$. 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)

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

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; Spohn 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

Similarly to Eichhorn et al. (2018), we will implement non-monotonic inferences by plausibility relations on possible worlds by instantiating preferential models (Makinson 1994) with total preorders resp. Ordinal Conditional Functions (OCF, (Spohn 1988; Spohn 2012)) which we derive from the statistical data of experiments via inference patterns.

4.1 Preferential Inference

For non-monotonic 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 $B \preccurlyeq A$. Non-monotonic 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: $\mathit{Bel}(\Psi) = \mathit{Th}(\min(\preccurlyeq_\Psi))$, 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 non-monotonic 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 'Balloon' content from Singmann et al.'s (2016) experiments – "If a balloon is pricked with a needle, then it will pop". Let N indicate that a balloon is pricked with a needle (n), or not ($\overline{n}$), and P indicate that the balloon has popped (p), or not ($\overline{p}$). Here the possible worlds are $\{np, n\overline{p}, \overline{n}p, \overline{n}\,\overline{p}\}$. We define* *the epistemic state Ψ to be represented by the preorder*

$$np \approx_\Psi \overline{n}\,\overline{p} \prec_\Psi n\overline{p} \approx_\Psi \overline{n}p.$$

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

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. Also called a *ranking function*, an Ordinal Conditional Function (OCF, (Spohn 1988; Spohn 2012)) 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. Given a normalization constraint, 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 $A \mid\!\sim_\kappa B$) 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 $\kappa(np) = \kappa(\overline{n}\,\overline{p}) = 0$, $\kappa(n\overline{p}) = \kappa(n\overline{p}) = 1$. With κ we have $\kappa(np) < \kappa(n\overline{p})$, and thus $\kappa \models (p|n)$ and also $n \mid\!\sim_\kappa p$.*

5 Inference Patterns

Eichhorn et al. (2018) proposed an approach to combine all four inference rules into tuples called *inference patterns* in order to classify psychological findings. Their initial point are the inference rules and their respective inferences as shown in Table 4, followed by a formalization of what it means that it is *plausible* to draw conclusions according to these rules, as (re-)introduced in the following.

Definition 1 (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., ¬MP) in an inference scenario. The set of all 16 inference patterns is called $\mathcal{R}$.*

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 $A \mid\!\sim B$ is drawn. This is the case if and only if the worlds are ordered such 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 5 gives all of the plausibility constraints which are equivalent to using the inference rules.

Table 4: Overview of the inferences drawn or not drawn from "From A it (usually) follows that B" with respect to application of the inference rules.

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

Table 5: Constraints on the plausibility relation on worlds in order to satisfy inference rules.

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

To satisfy an inference pattern, the plausibility relation has to satisfy each of the constraints given in Table 5. 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 $(\text{MP}, \text{MT}, \neg\text{AC}, \text{DA})$ (which occurs as an individual pattern in the balloon example, see Table 9), the worlds have to be ordered such that all four constraints given in Table 6 are satisfied.

If for a given pattern ϱ, there is a plausibility relation $\preccurlyeq$ that satisfies $\mathcal{C}(\varrho)$, that is, there is a total preorder on the worlds which is in accordance with plausible reasoning, ϱ can be deemed to be rational. Therefore, we call an inference pattern *rational* if and only if there is a plausibility relation $\preceq$ that satisfies the inference pattern. Note that, similar to more classical approaches, rationality is understood in terms of compliance with logic. However, here we use non-monotonic logics and its preferential models as norms for rational reasoning behavior.

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 $\overline{A}\,\overline{B} \prec A\overline{B} \preccurlyeq AB \prec \overline{A}B \preccurlyeq \overline{A}\,\overline{B}$.

Eichhorn et al. (2018) used this approach to analyze the combination of inference rules in an experiment. We will perform a similar analysis on the experimental data presented in Section 2, however, since now we are dealing with probabilistic endorsements, in order to enable such analysis,

Table 6: Constraints for the inference pattern (MP, MT, ¬AC, DA).

	$\{AB \prec A\overline{B}, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B \preccurlyeq AB, \overline{A}\,\overline{B} \prec \overline{A}B\}$
yields	$\overline{A}\,\overline{B} \prec \overline{A}B \preccurlyeq AB \prec A\overline{B}$

we will propose an enhancement of the inference patterns in the following section.

6 Inference Patterns with Endorsement

A probabilistic endorsement of a conclusion describes the degree of subjective belief an individual has in that world, in the range 0-100%. We consider endorsements that are $\geq 50\%$ as $true$, i.e., the inference form has been applied and otherwise $false$– the inference form has not been applied.

Definition 2. *For a conditional $(B|A)$ and an ordinal conditional function κ, we say κ accepts $(B|A)$ with strength s if $\kappa \models (B|A)$ and $s = \kappa(A\overline{B}) - \kappa(AB)$. We call $s = s_\kappa((B|A))$ the* κ-strength *of $(B|A)$.*

Definition 3. *An inference rule r is more* endorsed *than an inference rule r' with respect to a ranking function κ if $s_\kappa(\varphi_r) > s_\kappa(\varphi_{r'})$ holds for the associated conditionals $\varphi_r, \varphi_{r'}$.*

Following Definition 3, we examined the relationships between endorsements in Singmann et al.'s (2016) experimental data on both aggregate *and* individual level. We also specified a *difference tolerance* of 5, meaning that two endorsements (in the range 0-100%) will be considered as equal if the difference between them is $\leq 5\%$.

Example 3 (Ranking of endorsements)**.** *Let us consider the task 'Girl' in reduced inference (Table 3). Given the endorsements MP: 32%, MT: 44%, AC: 88%, DA: 85%, the respective rankings would then be ¬MP ≻ ¬ MT (false, not applied) and AC = DA (true, applied). Additionally, the corresponding (non-enhanced) inference pattern would be $(\neg MP, \neg MT, AC, DA)$.*

Note that for negated inference rules, the preorder $\succeq$ is reversed, i.e., in the example above, both MP and MT are not applied (i.e. false), but the endorsement of MP is lower than MT. This results in MP ≻ MT.

The derived rankings of the average inference form endorsements are presented in Table 7. With them, we can now *enhance* the inference patterns from Eichhorn et al. (2018) by statements about the strengths of inference rules.

Definition 4 (Extended Inference Pattern)**.** *An* extended 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., ¬MP) in an inference scenario, possibly together with statements about the ranking of endorsements of these (negated) inference rules.*

Example 4. *In Ex. 3, we obtain the extended inference pattern $(\neg MP, \neg MT, AC, DA; \neg MP \succ \neg MT, AC = DA)$. It shows that MP and MT have not applied, AC and DA have been applied, MP is* less *endorsed than MT, and AC and DA are endorsed equally.*

7 Explaining Human Inferences

In the introduction we have briefly introduced the theory of mental models (Johnson-Laird and Byrne 2002). This theory argues that people represent possibilities (we call them here possible worlds) that can depend on "knowledge, pragmatics, and semantics". As this theory can represent even the case $A\overline{B}$ the question arises, which worlds are preferred over others. The state of art in psychological research implicitly suggests that there are some orders on worlds.

Another psychological experiment (Barrouillet, Grosset, and Lecas 2000) suggests the order $AB \prec \overline{A}\,\overline{B} \prec \overline{A}B$. This has been so far identified experimentally only on the aggregate level (i.e., the mean of answers), but it has not yet been shown if this order holds for the individual reasoner. This is, however, most important as modeling each individual is the preferred goal of cognitive modeling, since models for the aggregate can distort theories (Fisher, Medaglia, and Jeronimus 2018). In the following, we introduce the necessary definitions and analyses to support or reject the claimed order and to analyze the inference patterns.

Definition 5. *A* preferential mental model *is a set of possible worlds together with a total preorder.*

As Eichhorn et al. (2018) explained, inference patterns can be realized by preferential mental models. Now, together with the endorsements, we are able to refine these preferential mental models. For that, we make use of ordinal conditional functions to be able to use arithmetics for the comparisons. However, in order to only make use of arithmetics on an intuitive level, we restrict the exploitation of these comparisons to basic cases. For instance, via ordinal conditional functions, the statement MP ≻ MT translates into $\kappa(A\overline{B}) - \kappa(AB) > \kappa(A\overline{B}) - \kappa(\overline{A}\,\overline{B})$, which is equivalent to $\kappa(AB) < \kappa(\overline{A}\,\overline{B})$. Note that $\kappa(A\overline{B})$ occurs in both differences which allows for an easy comparison by basic arithmetics. In this way, we obtain the following results for qualitative comparisons among the endorsements of inference rules:

$MP \succ MT$	$AB \prec \overline{A}\,\overline{B}$
$MP \succ AC$	$\overline{A}B \prec A\overline{B}$
$MT \succ DA$	$\overline{A}B \prec A\overline{B}$
$AC \succ DA$	$AB \prec \overline{A}\,\overline{B}$

Regarding disablers and alternatives, we translate their influence on the acceptance/endorsement of inference forms into these schemata, so that we are able to identify them in the preferential mental models. The presence of many disablers reduces the degree of belief in the logically valid MP and MT, and alternatives reduce the endorsement of AC and DA (Byrne 1989; Singmann, Klauer, and Beller 2016).

- Few disablers make the antecedent very informative for the consequent, similarly as in classical implications. Therefore, the logically valid MP and MT inference rules should be strong. So, we characterize this scenario by $MP \succeq AC$, which results in $\overline{A}B \preceq A\overline{B}$. Note that $MT \succeq DA$ yields the same constraint.
- Consequently, many disablers are modeled by $A\overline{B} \prec \overline{A}B$.

Table 7: Derived rankings of the average inference form endorsements for each task in each conditional presentation form, corresponding preferential mental models and scenarios. Scenarios that coincide with expected scenarios are marked in bold. In the rankings, inference forms that are True (applied, endorsement ≥ 50) are preceded with a 'T', ones that are False (not applied, endorsement < 50) with a 'F'. The average values of the inference form endorsements are presented in Table 3. ('Red.' - Reduced Inference, 'Full' - Full Inference, 'Dis' - Disablers, 'Alt' - Alternatives, 'Sc.' - Scenario)

Form	Task	Dis / Alt	Ranking of Endorsements	Preferential Mental Model	Sc. (Dis / Alt)
Red.	Predator	Few/Few	T: $MP \succ AC = MT \succ DA$	$AB \prec \overline{A}\,\overline{B} \prec \overline{A}B \prec A\overline{B}$	**Few/Few**
	Balloon	Few/Many	T: $MP \succ MT \succ DA \succ AC$	(no preferential mental model)	
	Girl	Many/Few	T: $AC = DA$; F: $MP \succ MT$	$A\overline{B} \preceq \overline{A}\,\overline{B} \prec AB \prec \overline{A}B$	Many/Many
	Coke	Many/Many	T: $MP = MT \succ DA \succ AC$	$\overline{A}\,\overline{B} \prec AB \prec \overline{A}B \prec A\overline{B}$	Few/Many
Full	Predator	F / F	T: $MP = AC \succ DA$ $AC = MT; MP \succ MT$	$AB \succ \overline{A}\,\overline{B} \succ A\overline{B}, \overline{A}B$'	**Few/Few**
	Balloon	Few/Many	T: $MP \succ MT \succ AC = DA$	$AB \prec \overline{A}\,\overline{B} \prec \overline{A}B \prec A\overline{B}$	Few/Few
	Girl	Many/Few	T: $AC = DA \succ MP = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B} \prec \overline{A}B$	**Many/Few**
	Coke	Many/Many	T: $MP = MT \succ AC = DA$	$AB, \overline{A}\,\overline{B} \prec \overline{A}B \prec A\overline{B}$	Few/Few

Table 8: Number of individuals out of 199 that have the same ranking that is found on the aggregate level (shown in Table 7).

Task	# Individuals Reduced	# Individuals Full
Predator	3 (1.51%)	0 (0.0%)
Balloon	3 (1.51%)	5 (2.51%)
Girl	27 (13.57%)	11 (5.53%)
Coke	0 (0.0%)	4 (2.01%)

- Few alternatives make the antecedent very plausible when observing the consequent, so particularly AC should be strong. We model this via $AC \succeq DA$ which gives us $AB \preceq \overline{A}\,\overline{B}$.
- Consequently, many alternatives are modeled by $\overline{A}\,\overline{B} \prec AB$.

Please keep in mind that artifacts that contradict these schematic classifications may arise due to the general plausibility of A and B in the background knowledge of the individuals.

Aside from the inference patterns, Table 7 also presents the corresponding preferential mental models and the respective scenarios.

If we look at the aggregate case (Table 7) for the reduced inference presentation form, the scenarios respective to the induced preferential mental models do not necessarily correspond to the original quantification of disablers and alternatives associated with the tasks' contents. E.g., the 'Coke' task, leads to the question whether alternatives are more influential than disablers when a content has 'Many' of both associated with it. Additionally, when looking into the 'Balloon' task, the inference pattern derived from aggregate data is inconsistent, i.e. it induced no preferential mental models. This may happen due to too divergent views of the individuals. In their analysis, Singmann et al. (2016) showed that when presented with a reduced inference, individuals tend to rely more on their background knowledge and have a stronger influence by the corresponding disablers and alternatives, in contrast to the full inference. That effect can also be seen here, as in the full inference presentation form, the scenarios identify 'Few' disablers and alternatives even when there are 'Many', meaning that individuals were not integrating their background knowledge as much. An exception is the 'Girl' content, where disablers seem to be exceptionally influential, which is also visible when looking at the average inference forms endorsements (Table 3).

However, to which extent do these findings on the aggregate level hold for the individual reasoner? We looked into each participant's endorsements and found out for how many participants the ranking derived from the aggregate data holds. The numbers are shown in Table 8. We can immediately see that the number of individuals that are captured by the aggregate rankings is exceptionally low. This was an expected outcome (Fisher, Medaglia, and Jeronimus 2018), and strongly confirms our need for individual analysis. Therefore, we also derived the rankings, the induced preferential models and scenarios for each individual separately.

The different rankings found for at least 5% of participants are shown in Table 9. It can be seen that individuals are not divided in only a few largely populated groups, but there are multiple different rankings found across the 199 participants. For each pattern the size of the participant group is also presented in the table. In most cases, the scenarios identify the same quantities of disablers and alternatives as originally associated with the respective tasks. This is extremely important, as we can see that groups of individuals *do* interpret the conditionals as intended. If we only focused on the aggregate analysis, we would most likely dismiss the tasks in the reduced inference case due to the lack of correspondence between derived scenarios and expected ones.

Additionally, the inference patterns whose preferential mental models do not identify the same quantities are still present among a larger group of participants, which points to possible differences in conditional interpretation and different knowledge bases. E.g., the third most frequent pattern

Table 9: Individual rankings, preferential mental models and corresponding scenarios for all contents in both conditional presentation forms. The rankings and preferential models are listed in descending order of the frequencies of the appertaining extended inference patterns, i.e., the first line corresponds to the most frequent extended inference pattern. Only inference patterns that were found for at least 5% of participants are taken into consideration and the exact number of individuals for each ranking is presented. Scenarios that correspond to the expected scenarios are marked in bold. In the rankings, inference forms that are True (applied, endorsement ≥ 50) are preceded with a 'T', ones that are False (not applied, endorsement < 50) with a 'F'. ('Red.' - Reduced Inference, 'Full' - Full Inference, 'Dis' - Disablers, 'Alt' - Alternatives, 'Ind.' - Individuals)

Task	Dis / Alt	Form	Ranking	Preferential Mental Model	Scenario (Dis / Alt)	# Ind.
Predator	Few/Few	Red. Full	T: $MP = AC = DA = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B$	**Few/Few**	49 59
Balloon	Few/Many	Red.	T: $DA = MP = MT \succ AC$	$\overline{A}\,\overline{B} \prec AB \prec A\overline{B}, \overline{A}B$	**Few/Many**	16
			T: $MP = MT \succ DA$; F: AC	$\overline{AB} \prec \overline{A}B \preceq AB \prec A\overline{B}$	**Few/Many**	15
			T: $MP = AC = DA = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B$	Few/Few	12
			T: $MP = MT \succ DA \succ AC$	$\overline{A}\,\overline{B} \prec AB \prec \overline{A}B \prec A\overline{B}$	**Few/Many**	10
		Full	T: $MP = AC = DA = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B$	Few/Few	46
			T: $MP = MT \succ DA$; F: AC	$\overline{A}\,\overline{B} \prec \overline{A}B \preceq AB \prec A\overline{B}$	**Few/Many**	13
			T: $MP = MT \succ AC = DA$	$AB, \overline{A}\,\overline{B} \prec \overline{A}B \prec A\overline{B}$	**Few/Many**	13
			T: $DA = MP = MT \succ AC$	$\overline{A}\,\overline{B} \prec AB \prec A\overline{B}, \overline{A}B$	**Few/Many**	12
Girl	Many/Few	Red.	T: $AC = DA$; F: $MP \succ MT$	$A\overline{B} \preceq \overline{A}\,\overline{B} \prec AB \prec \overline{A}B$	Many/Many	27
			T: $AC = DA$; F: $MP = MT$	$A\overline{B} \preceq AB, \overline{A}\,\overline{B} \prec \overline{A}B$	**Many/Few**	19
			T: $AC = DA \succ MT$; F: MP	$\overline{A}\,\overline{B} \prec A\overline{B} \preceq AB \prec \overline{A}B$	Many/Many	17
			T: $AC = DA$; F: $MT \succ MP$	$A\overline{B} \preceq AB \prec \overline{A}\,\overline{B} \prec \overline{A}B$	**Many/Few**	15
			T: $AC = DA \succ MP$; F: MT	$AB \prec A\overline{B} \preceq \overline{A}\,\overline{B} \preceq \overline{A}B$	**Many/Few**	12
			T: $AC = DA \succ MP = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B} \prec \overline{A}B$	**Many/Few**	11
		Full	T: $MP = AC = DA = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B$	Few/Few	29
			T: $AC = DA$; F: $MP = MT$	$A\overline{B} \preceq AB, \overline{A}\,\overline{B} \prec \overline{A}B$	**Many/Few**	13
			T: $AC = DA \succ MP = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B} \prec \overline{A}B$	**Many/Few**	11
			T: $AC = DA$; F: $MP \succ MT$	$A\overline{B} \preceq \overline{A}\,\overline{B} \prec AB \prec \overline{A}B$	Many/Many	10
			T: $AC = DA \succ MP$; F: MT	$AB \prec A\overline{B} \preceq \overline{A}\,\overline{B} \preceq \overline{A}B$	**Many/Few**	10
			T: $AC = DA \succ MT$; F: MP	$\overline{A}\,\overline{B} \prec A\overline{B} \preceq AB \prec \overline{A}B$	Many/Many	10
Coke	Many/Many	Red.	(no consistent ranking found)	–	–	–
		Full	T: $MP = AC = DA = MT$	$AB, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B$	Few/Few	23
			T: $MP = MT = AC \succ DA$	$AB, \overline{A}\,\overline{B} \prec \overline{A}B \prec A\overline{B}$	Few/Few	11

for the 'Balloon' content in the reduced inference case does not suggest that those individuals were able to incorporate 'Many' alternatives when reasoning.

When comparing reduced inference with full inference, it can be seen that patterns that induce a 'Few/Few' scenario even if there are 'Many' disablers and alternatives present are rather frequent, meaning that the influence of background knowledge has been suppressed. However, there is still a significant amount of individuals who nevertheless successfully integrate information about the disablers/alternatives when reasoning. A conclusion about the influence of the full inference in contrast to the reduced one on the effect of disablers and alternatives should absolutely *not* be derived based on aggregate data. As we show here, individuals and their interpretations *differ*.

It is also noteworthy that even though there are different inference form endorsement combinations present among individuals, after deriving the corresponding preferential mental models, they all suggest the same interpretation. The presence of a certain amount of disablers and alternatives can be modeled and endorsed in various ways by different people.

8 Discussion and Conclusion

We followed the inference evaluation approach by applying logic based on conditionals and plausible reasoning proposed by Eichhorn et al. (2018) and extended it towards *probabilistic endorsements*. We do not only take into consideration whether an inference form has been applied or not, but also look into the relationships between the subjective degrees of belief in said inference forms for various contents. Using OCFs (Spohn 2012) a plausibility relation on possible worlds was defined in order to obtain a preferential entailment.

The beauty of this interdisciplinary field is that different formalisms can be used in analysis and with that we can get insight into human reasoning from many different perspectives which can be joined to get an even better understanding of reasoning processes. Given the preferential character of probabilistic endorsements, ranks are a natural approach to consider. Our contribution is to show how ranking functions can be applied to probabilistic conditional reasoning experimental data and what we can learn from them.

The extended inference patterns reveal in an abstract way how people *understood* the task they were given. Our focus was on tasks with a varying amount of disablers and alternatives – events that make humans diverge from logical reasoning. They are especially influential in a reduced inference presentation form, when individuals are not bound by a conditional rule but can rather integrate their personal background knowledge on a higher level. Our approach is flexible enough to be able to show the impact of disablers and alternatives. We are already familiar with the fact that humans deviate from classical logic when reasoning, so shifting the focus of research to everyday contents is important. Understanding how background knowledge, personal and cultural differences influence reasoning is of our interest.

As illustrated by the large variety in the derived inference patterns we can see the effect of individual differences (e.g. substantial cultural differences) and how human reasoning can be very diverse. Therefore, an aggregate analysis approach might not always be appropriate to get a better understanding of inference mechanisms, but the individual differences play a big role and should be taken into consideration.

Additionally, by performing individual analysis we can also learn whether certain experimental content managed to achieve its goal. For example, the 'Coke' content is supposed to have 'Many' disablers and alternatives associated with it, which as true as it might be, it does not seem to be understood that way by the participants. In Table 9 we see that in the reduced inference case, where the background knowledge should be dominating, no consistent pattern was found. That means that there was not a single group of individuals formed by at least 5% of the total participants that understood the task in the same way, which points to the need for reconsideration of the chosen content.

The inference pattern derived from aggregate data for the 'Balloon' task in the reduced inference case is inconsistent. Contradictory patterns show irrationality when found on an individual level, e.g. the three individual participants that had the same endorsement ranking reasoned irrationally, which is, of course, a common occasional human trait, and our approach can account for this! However, having found an inconsistent pattern on the aggregate level means that the individuals' perspectives and interpretation are diverging too much in order to be aggregated consistently. This supports the idea that an *individual* analysis approach is necessary. Moreover, it also indicates a potential requirement to reconsider whether such content is suitable to test human reasoning. Naturally, in order to determine proper task contents a significant increase in various experimental data is required.

To conclude, we analyzed the same experimental data on two levels – aggregate and individual. In many ways we showed that the focus undoubtedly needs to be switched to the *individual*. Humans are diverse, their personal experiences lead to diverging background knowledge and interpretation abilities. In order to make a larger leap forward towards understanding the human reasoning processes, these differences have to be taken into consideration and modeling approaches should be able to account for the deviations between individuals. We found preferential mental models on the individual level whose interpretations give us insight into how the experimental content manipulation affected (or not) the participants' reasoning. Generally, the fact that many different individual inference patterns induce the same mental models points to a significant strength of our approach – using mental models is more fundamental than representation forms that heavily rely on statistics.

The next step in this work would be to look more into the specifics of the relevant background knowledge and its influence on reasoning. Additionally, an even larger focus on the individual would be of interest. For instance, how do the individual's subjective believes and inference mechanism change between different contents or conditional presentation forms? The theoretical foundation of our approach would allow for gaining more insight into such questions and aid in getting a better understanding of the individual reasoning processes.

References

Barrouillet, P.; Grosset, N.; and Lecas, J.-F. 2000. Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition* 75(3):237–266.

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

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.

Eichhorn, C.; Kern-Isberner, G.; and Ragni, M. 2018. Rational inference patterns based on conditional logic. In *AAAI Conference on Artificial Intelligence*, volume 32, 1827–1834.

Fisher, A. J.; Medaglia, J. D.; and Jeronimus, B. F. 2018. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences* 115(27):E6106–E6115.

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

Johnson-Laird, P. N., and Byrne, R. M. 1991. *Deduction*. Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., and Byrne, R. M. 2002. Conditionals: a theory of meaning, pragmatics, and inference. *Psychological review* 109(4):646.

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.

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.

Ragni, M.; Dames, H.; and Johnson-Laird, P. N. 2019. A meta-analysis of conditional reasoning. In Stewart, T. C., ed., *Proceedings of the 17th International Conference on Cognitive Modeling*, 151–156. Waterloo, Canada: University of Waterloo: https://iccm-conference.neocities.org/2019/proceedings/index.html.

Ragni, M.; Kola, I.; and Johnson-Laird, P. N. 2018. On selecting evidence to test hypotheses: a theory of selection tasks. *Psychological Bulletin* 144(8):779–796.

Singmann, H.; Klauer, K. C.; and Beller, S. 2016. Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology* 88:61.

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.