Probabilistic Conditional Reasoning: Disentangling Form and Content with the Dual-Source Model

Henrik Singmann and Karl Christoph Klauer
Albert-Ludwigs-Universität Freiburg

Sieghard Beller
University of Bergen

To appear in: *Cognitive Psychology*

The present research examines descriptive models of probabilistic conditional reasoning, that is of reasoning from uncertain conditionals with contents about which reasoners have rich background knowledge. According to our dual-source model, two types of information shape such reasoning: knowledge-based information elicited by the contents of the material and content-independent information derived from the form of inferences. Two experiments implemented manipulations that selectively influenced the model parameters for the knowledge-based information, the relative weight given to form-based versus knowledge-based information, and the parameters for the form-based information, validating the psychological interpretation of these parameters. We apply the model to classical suppression effects dissecting them into effects on background knowledge and effects on form-based processes (Exp. 3) and we use it to reanalyse previous studies manipulating reasoning instructions. In a model-comparison exercise, based on data of seven studies, the dual-source model outperformed three Bayesian competitor models. Overall, our results support the view that people make use of background knowledge in line with current Bayesian models, but they also suggest that the form of the conditional argument, irrespective of its content, plays a substantive, yet smaller, role.

Keywords: conditional reasoning; probabilistic reasoning; dual-source model; measurement model; meta-analysis

Introduction

It is difficult to overstate the influence Bayesian approaches have had on the development of theories in cognitive psychology in the last few decades. Across diverse domains – ranging from low-level phenomena such as perception to high-level phenomena such as argumentation – Bayesian models often provide an unprecedented level of explanatory power (Chater, Oaksford, Hahn, & Heit, 2010). A core assumption of such models is that subjective degrees of belief can be modeled as probabilities obeying the axioms of probability theory.

In the field of reasoning, people evaluating an argument have traditionally been asked to assume that the stated premises hold true and to ignore any background knowledge elicited by the contents of the premises. In tune with these instructions, theoretical accounts have often assumed that reasoning is performed on relatively abstract representations of the argument form (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Rips, 1994). The advent of Bayesian and related probabilistic approaches has led to what has been called the *new paradigm* (Over, 2009). In the new paradigm, the reasoning problems are typically couched in everyday, real-world contents, and reasoners are not instructed to disregard what they know about the contents.

Using such contents, it turned out that it is easy to construct arguments that are logically valid, but yield a conclusion that most people reject as highly improbable. Conversely, it was found easy to construct arguments that are logically invalid, but yield a conclusion that most people accept as highly probable (in the conext of conditional reasoning see Nickerson, 2015, chap. 14; Byrne, 1991; Singmann & Klauer, 2011). These and related findings have led many theorists to make a strong case that human reasoning is guided not by logic, but by probability. For example, Chater and Oaksford (2001) state that "we see probability theory as a

Henrik Singmann and Karl Christoph Klauer, Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany. Sieghard Beller, Department of Psychosocial Science, University of Bergen, Bergen, Norway. Henrik Singmann is now at the Department of Psychology, University of Zürich, Switzerland.
The first two authors contributed equally to this work.
This work was supported by Grant KL 614/33-1 to Karl Christoph Klauer and Sieghard Beller from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program "New Frameworks of Rationality" (SPP 1516). We thank the members of the SPP 1516 for their ideas and discussions about this work.
Correspondence concerning this article should be addressed to Henrik Singmann (singmann@psychologie.uzh.ch).

wholesale replacement for logic as a computational level theory of what inferences people should draw." (p. 208). Prominent Bayesian models assume that reasoning amounts to the assessment of probabilities of conclusions based on what the reasoners know about the contents of conclusions and premises; reasoning is thus conceptualized as probabilistic and content-driven (e.g., Baratgin & Politzer, 2006; Cruz, Baratgin, Oaksford, & Over, 2015; Oaksford & Chater, 2007; Pfeifer & Kleiter, 2010).

The focus of the current work is on a model that integrates these two seemingly irreconcilable positions within the new paradigm. We will show that there are content-independent effects of different argument forms that are not adequately captured by Bayesian models. Hence, we propose that reasoning is influenced by two different and independent cognitive processes – a probabilistic process in line with extant Bayesian models, which we call *knowledge-based*, and a content-independent process driven by the form of the argument, which we call *form-based*. In this view, reasoners' evaluations actually reflect a mixture of form-based and knowledge-based information. These assumptions are explicated formally in our *dual-source model* (DSM; Klauer, Beller, & Hütter, 2010) elaborated on below.

Probabilistic Conditional Reasoning

A conditional rule links two propositions, an antecedent p and a consequent q, in the form "if p then q". Inference tasks in conditional reasoning (for an overview, see Nickerson, 2015) typically present the conditional rule as major premise and one of p, q, or their negations as minor premise. Reasoners are asked to assess a proposed conclusion on the basis of this information. According to classical logic, two of the usually studied inferences are valid; the truth of the premises entails the truth of the conclusion:

> Modus ponens (MP): Given "if p then q" and "p", it follows that "q".
>
> Modus tollens (MT): Given "if p then q" and "not q", it follows that "not p".

The two so-called reasoning fallacies are invalid; the truth of the premises does not entail the truth of the conclusion:

> Affirmation of the consequent (AC): Given "if p then q" and "q", it follows that "p".
>
> Denial of the antecedent (DA): Given "if p then q" and "not p", it follows that "not q".

Studies in the old paradigm typically employed contents for which participants have little prior knowledge (e.g., a major premise might be: "If there is a vowel on the blackboard then there is an even number on the blackboard"), and participants are asked to evaluate the logical validity of the above conditional arguments. Naive reasoners almost unanimously accept MP as valid, whereas acceptance rates for MT are significantly smaller, although typically still above 50%. AC and DA are often erroneously accepted as valid with acceptance rates of AC sometimes reaching those of MT (e.g., Evans, 1993; Schroyens & Schaeken, 2003).

In contrast, studies in the new paradigm usually employ everyday contents (e.g., a premise might be: "If a balloon is pricked with a needle then it will pop"), and participants are to evaluate the probability or plausibility of the conclusions of the different conditional arguments on a graded scale. In what follows, we will use the term *endorsement* to refer to these graded responses.

Antecedent and consequent of so-called *causal* conditionals are at least weakly related as cause and effect. For causal conditionals, content is usually characterized and sometimes manipulated in terms of two different types of counterexamples: *disablers* that prevent the consequent in the presence of the antecedent and *alternatives* that bring about the consequent in the absence of the antecedent. For instance, for the conditional "If a person drinks a lot of coke then the person will gain weight", heavy exercising is a disabler that prevents weight gain, whereas eating food rich in calories is an alternative cause of gaining weight. A set of seminal experiments by Cummins (1995; Cummins, Lubart, Alksnis, & Rist, 1991; see also Thompson, 1994) established that the availability or likelihood (Geiger & Oberauer, 2007) of disablers is negatively related to the endorsement of the formally valid inferences MP and MT, whereas the availability or likelihood of alternatives is negatively related to the endorsement of the reasoning fallacies AC and DA (for reviews see Beller & Kuhnmünch, 2007; Politzer, 2003).

Oaksford, Chater, and Larkin's (2000) model of probabilistic conditional reasoning. An influential Bayesian model of probabilistic conditional reasoning was proposed by Oaksford et al. (2000). The model assumes that reasoning amounts to assessing probabilities of conclusions based on one's background knowledge. More precisely, when asked to evaluate an inference such as MP, "Given '*If p then q*' and p, how likely is q?", individuals consult their background knowledge regarding p and q and assess the conditional probability of the conclusion q given minor premise p. Thus, endorsement E is modeled as $\mathrm{E(MP)} = P(q|p)$. The joint probability distribution of p, q, and their negations $\neg p$, $\neg q$ can be parameterized in terms of three parameters, $a = P(p)$, $b = P(q)$, and $e = P(\neg q|p)$ as shown in Table 1, which leads to the following model predictions:

Table 1
Joint probability distribution for a conditional "If p then q".

	q	$\neg q$
p	$a(1-e)$	ae
$\neg p$	$b-a(1-e)$	$(1-b)-ae$

Note. $a = P(p)$, $b = P(q)$, and $e = P(\neg q|p)$.

$$\mathrm{E(MP)} = P(q|p) = (1-e) \tag{1}$$

$$\mathrm{E(MT)} = P(\neg p|\neg q) = \frac{1-b-ae}{1-b} \tag{2}$$

$$\mathrm{E(AC)} = P(p|q) = \frac{a(1-e)}{b} \tag{3}$$

$$\mathrm{E(DA)} = P(\neg q|\neg p) = \frac{1-b-ae}{1-a} \tag{4}$$

The endorsements of the four inferences provide four independent data points, whereas there are only three free model parameters, a, b, and e. In consequence, Oaksford et al.'s (2000) model can be fitted to data and the differences between observed and predicted values indicate whether or not the reasoners' responses are consistent with the axioms of probability theory as postulated by the model.[1]

Fitting the model to empirical data, the Bayesian model outperformed a classical logic-based model for data obtained in the old paradigm (Oaksford & Chater, 2003) and described data from studies in the new paradigm well (Oaksford et al., 2000). Moreover, there was evidence for the intended interpretation of the model parameters a and b in terms of $P(p)$ and $P(q)$, respectively: The parameter estimates of a (b) were small in value when pretesting showed that antecedent cases p (consequent cases q) were rare compared to when they were frequent.

Disentangling form and content: The dual-source model. In the old paradigm, a focus on logical form was encouraged through the use of contents for which little background knowledge is available (such as arbitrary rules obeyed by letters on a blackboard), leading to a neglect of the potential impact of content-related variables, which were usually held constant. In the new paradigm, meaningful contents are used for which background knowledge is available, and content-related variables such as the availability of different kinds of counterexamples are manipulated, but there is now a neglect of the potential impact of logical form.

For example, Oaksford et al.'s (2000) model refers only to the joint probability distribution shown in Table 1, but it does not specify a role for the conditional rule as such. In fact, a conditional inference such as MP:

- If a balloon is pricked with a needle, then it will pop. A balloon is pricked with a needle. How likely is it that it will pop?

can be meaningfully evaluated even if the conditional premise is left out and thus in the absence of a definite logical form:

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

Simply mentioning p ("a balloon is pricked with a needle") and q ("the balloon pops") is sufficient to elicit the background knowledge summarized in the joint probability distribution of Table 1 which in turn allows one to assess the conditional probability of q given p. This makes it difficult to disentangle whether the logical MP form as such contributes anything over and above the content-driven probabilistic assessment.

The present dual-source model (DSM) builds on the dual-source framework (Beller & Spada, 2003; Klauer et al., 2010) according to which both logical form *and* content shape reasoning with realistic materials. Reasoners' responses are seen as reflecting a weighted integration of both sources of information, logical form and content.

To disentangle the impacts of logical form and content, the DSM contrasts (a) responses to the conditional inferences and (b) responses to reduced inferences that leave out the conditional rule. As just exemplified for the MP form, the reduced MP inference presents only the minor premise p and asks for an assessment of the conclusion q. This yields a baseline that reflects only the content-related contribution (Liu, 2003, see also Beller, 2008; Beller & Kuhnmünch, 2007).

A few studies have contrasted reduced and full inferences and found that conclusion endorsement increased when the conditional was present as compared to when the conditional was absent. This increase was relatively content-independent and especially pronounced for MP and MT (Klauer et al., 2010; Liu, 2003; Matarazzo & Baldassarre, 2008).

According to the DSM, responses to the reduced inferences are probabilistic assessments based on the joint probability distribution of Table 1 like in Oaksford et al.'s (2000) model. Adding the conditional rule results in a definite logical form. According to the DSM, reasoners also assess the logical form where present resulting in a form-based response proposal, and observed responses then reflect a weighted mixture of both knowledge-based and form-based response proposals (Klauer et al., 2010).

Let us refer to the content domain addressed in the four inferences with a given conditional by C (e.g., the domain of the behaviors of balloons pricked or not pricked with a needle) and to the four inferences by x, $x \in \{$MP, MT, AC,

[1]Deriving the likelihood function for the data is not trivial as responses are restricted to the probability scale. Oaksford et al. (2000; Oaksford & Chater, 2007) fit the data by minimizing the sum of squared deviations of model predictions and data, a solution also adapted by Klauer et al. (2010) and in the present study.

Table 2
Parameters of the dual-source model.

Par.	Interpretation
τ	Degree to which an inference is seen as logically warranted
ξ	Knowledge-based response proposal (based on the parameters given in Table 1)
λ	Relative weight given to form-based versus knowledge-based evidence

DA}. We continue to refer to the reduced inferences as MP, MT, AC, or DA according to the logical form that results if the rule is added.

Endorsement of the reduced problem x (e.g., MP) with content C, $\mathrm{E}_r(C, x)$, reflects the conditional probability of the conclusion (e.g., q) given the premise (e.g., p) just like in Oaksford et al.'s (2000) model. We refer to the conditional probability by the DSM's knowledge parameter $\xi(C, x)$:

$$\mathrm{E}_r(C, x) = \xi(C, x). \quad (5)$$

Corresponding to the four inferences x, there are four knowledge parameters ξ per content which are expressed as in Oaksford et al.'s (2000) model by three parameters a, b, and e per content (see Equations 1 to 4).

Endorsement of the full inference x, $\mathrm{E}_f(C, x)$, is a weighted average of a form-based component, multiplied by the weight λ given to the form-based information, and the knowledge-based component $\xi(C, x)$, multiplied by $(1 - \lambda)$, $0 \leq \lambda \leq 1$.

The form-based component reflects the subjective degree of belief $\tau(x)$ in the validity of the (full) inference x on a probability scale. When uncertain about the validity of the inference, with probability $1 - \tau(x)$, reasoners fall back on their background knowledge. Thus, the form-based component is itself a mixture, with mixture weights $\tau(x)$, given by $\tau(x)\times 1+(1-\tau(x))\times\xi(C, x)$. Taken together, the DSM predicts that responses to full probabilistic conditional inferences are given by

$$\mathrm{E}_f(C, x) = \lambda\{\tau(x) + (1 - \tau(x)) \times \xi(C, x)\} + (1 - \lambda)\xi(C, x), \quad (6)$$

where $0 \leq \lambda, \tau, \xi \leq 1$. An overview of the DSM parameters is given in Table 2.

The knowledge parameters can be uniquely estimated because they are identified from the observed endorsements of reduced problems (see Eq. 5). The parameters $\tau(x)$ and λ cannot be uniquely estimated on the basis of the observed endorsements to reduced and full problems; only their products $\lambda\tau(x)$ are identified (see the correction to Klauer et al., 2010). To fix the scale for these parameters, we set the largest of the $\tau(x)$ equal to one. This yields unique parameter estimates for λ and $\tau(x)$. Differences in the overall level of the profile of the $\tau(x)$ parameters over the four inferences are thereby represented in the λ parameter, whereas profile shape is reflected in the $\tau(x)$ parameters. The $\tau(x)$ parameters accordingly quantify the relative support for the respective conclusions elicited by the full inference forms x relative to each other. The λ parameters quantify the *overall* weight given to conditional rules and inferences based on them.

Figure 1 shows the profile of τ parameters estimated from previous experiments. It can be seen that it mirrors the pattern of acceptance rates with MP > MT > AC ≥ DA described above for reasoning based on materials for which little background knowledge is available.

The DSM's predictions for the basic paradigm. The model makes predictions for the pattern of observed endorsements for reduced and full inferences, $\mathrm{E}_r(C, x)$ and $\mathrm{E}_f(C, x)$. For the reduced inferences, these are the same as those of the Oaksford et al.'s (2000) model, namely that the endorsements are consistent with conditional probabilities from a joint probability distribution. As shown in the online appendix, this is equivalent to the constraint that for each con-

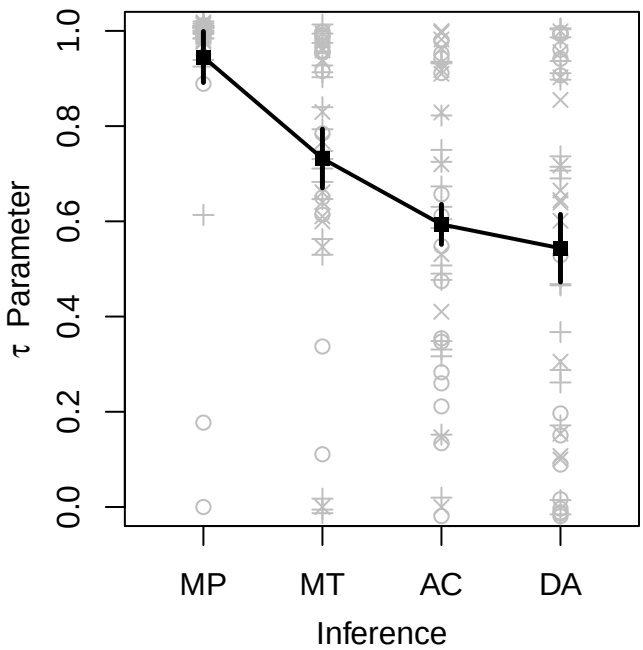

Figure 1. τ parameter from Experiments 1, 3 (condition "if-then"), and 4 of Klauer et al. (2010). Mean values are displayed in black, error bars show difference-adjusted 95% Cousineau-Morey-Baguley intervals (Baguley, 2012). Individual participants' values are displayed in grey (∘: Experiment 1; +: Experiment 3; ×: Experiment 4); in case of points with identical values a small amount of vertical jitter was added to avoid perfect overlap.

tent C,

$$(1 - \mathrm{E}_r(C, \mathrm{MP}))\mathrm{E}_r(C, \mathrm{MT})(1 - \mathrm{E}_r(C, \mathrm{DA}))\mathrm{E}_r(C, \mathrm{AC}) = \mathrm{E}_r(C, \mathrm{MP})(1 - \mathrm{E}_r(C, \mathrm{MT}))\mathrm{E}_r(C, \mathrm{DA})(1 - \mathrm{E}_r(C, \mathrm{AC})). \quad (7)$$

Turning to the full inferences, an additional prediction is that endorsement should increase, $\Delta(C, x) = \mathrm{E}_f(C, x) - \mathrm{E}_r(C, x) \geq 0$ for all C and x. As discussed above this is the pattern that is usually observed and we also generally find this in the new data presented here. An additional and unique prediction of the DSM is that for each inference x,

$$\frac{\Delta(C, x)}{1 - \mathrm{E}_r(C, x)} \text{ is not a function of } C. \quad (8)$$

As shown in the online appendix, observed endorsements are consistent with the DSM if and only if these predictions are met. We tested the non-parametric prediction of Equation 8 on data from 132 participants from six experiments who all used the same four contents C (see Table 3 below) in a meta-analytic fashion. This analysis (which is described in detail in the online appendix) revealed that C had no effect on the ratio confirming the prediction of Equation 8.[2] Taken together, for the basic paradigm the DSM does not only describe the usually obtained increase from reduced to full inferences, it also generates a new and unique prediction which is confirmed by the data.

Bayesian models of reduced and full inferences. Remember that reduced and full inferences differ by the conditional rule. How could adding a conditional rule be modeled in a Bayesian framework? Normatively, adding a premise (the conditional rule) implies updating the joint probability distribution of Table 1 using Bayes' rule. Updating amounts to computing the conditional probability distribution given the new premise. Unfortunately, Bayesian updating is defined only for premises that are events in the sample space of the probability distribution in question, which is not the case for conditional rules (at least as conceptualized in probabilistic conditional reasoning Nickerson, 2015, chaps. 9 and 10).

In consequence, there is no normative or agreed-upon solution of how to model updating by a conditional in a Bayesian framework. A few descriptive models have been proposed, however. For example, Oaksford and Chater (2007) state that the only effect of adding a conditional should be to reduce parameter e: "It seems that the only effect the assertion of the conditional premise could have is to provide additional evidence that q and p are related, which increases the assessment of $P_2(q|p)$ [i.e., of $1 - e$]" (p. 164). In this model, endorsement of the reduced and full problems is thus modeled as above (see Equations 1 to 4) in terms of three parameters a, b, and e per content C, but different parameters e and e' are used for the reduced and full problems, respectively, with $e' < e$. for each content with $e' < e$.

This version of Oaksford et al.'s (2000) model can be characterized by a different, unique prediction for the quantities $\frac{\Delta(C,x)}{1-\mathrm{E}_r(C,x)}$ that is, however, more complex than that shown in Equation 8 and correspondingly difficult to test directly. This model and two alternative Bayesian models that implement updating by conditionals are considered in the section "Goodness of fit Meta-Analysis" and compared to the DSM.

Scope of the Present Studies: Validating and Applying the DSM

Following the recommendations laid out by Heathcote, Brown, and Wagenmakers (2015) for good practices in cognitive modeling, we assess the DSM by evaluating model fit, by model comparisons, and by selective-influence studies. Because of the difficulties of directly testing complex nonlinear models' predictions for the data, a traditional approach is to fit such models to the data and to assess model fit (for an example in probabilistic reasoning, see Oaksford et al., 2000). A reasonable goodness of fit indicates that the ensemble of restrictions defining the model is satisfied to at least a reasonable degree of approximation. Moreover, comparing model fit across different models such as the DSM and the just-sketched descriptive Bayesian models reveals which set of restrictions describes the data better, the one defining the DSM or a set defining an alternative model.

Model parameters are mathematical entities that are as such devoid of psychological meaning. Selective-influence studies test whether the psychological interpretations attached to model parameters in the verbal description of the model are tenable. In selective-influence studies, manipulations are chosen that are expected – a priori and on theoretical grounds – to affect certain model parameters (such as the τ parameters) based on their intended psychological interpretation, while leaving other parameters unaffected. If the differences between the manipulated conditions do map on differences in the estimates of the targeted model parameters and leave the estimates of the other model parameters unaffected, the intended psychological interpretation of the model parameters in question is supported (Heathcote et al., 2015). As pointed out by Batchelder and Alexander (2013) in their discussion of selective-influence studies, "any paramet-

[2]Note that testing this prediction was not trivial as it involves a ratio. For 6.8% of the data this ratio was not identified as the endorsement for the reduced inferences, $\mathrm{E}_r(C, x)$, was 1 (we removed these data points from the analysis). Another problem was of numerical nature and jeopardized the statistical analysis. The ratio can take on extreme negative values if $\Delta(C, x)$ is negative and at the same time $\mathrm{E}_r(C, x)$ near 1. As noted above, $\Delta(C, x)$ is predicted to be positive which was the case for the vast majority of the data, but not always, as is to be expected in the presence of measurement error. We implemented different strategies of handling these outliers (including not removing any outliers) which all led to the same result, no effect of C.

Table 3
Contents used in Klauer et al.'s (2010) Experiments 1 and 3.

No.	Content	Disablers	Alternatives
1	If a predator is hungry then it will search for prey.	few	few
2	If a balloon is pricked with a needle then it will pop.	few	many
3	If a girl has sexual intercourse then she will be pregnant.	many	few
4	If a person drinks a lot of coke then the person will gain weight.	many	many

Note. Content is expressed in the form of a conditional "if p then q".

ric . . . model can be reparameterized in an unlimited number of ways, each yielding exactly the same model in terms of its ability to fit data. Many of these statistically equivalent models have parameters that have absolutely no psychological meaning, and, as a consequence, . . . validity [selective-influence] studies are essential. The usefulness of . . . models is that, if properly validated, they return measurements of latent cognitive processes, and this goes well beyond just providing statistically acceptable fits to the data" (p. 1209).

As an example, consider the knowledge parameters, $\xi(C, x)$. Klauer et al. (2010) used four contents, shown in Table 3, that systematically varied in the availability of disablers and alternatives. MP and MT (AC and DA) require that p is sufficient (necessary) for q to occur, and sufficiency (necessity) is questioned by disablers (alternatives). If our "knowledge parameters" ξ model such characteristics of reasoners' background knowledge then they should respond to variations between contents in the availability of counterexamples, of disablers and alternatives, in specific patterns. Klauer et al. (2010) indeed found that the knowledge parameters for MP and MT (for AC and DA) were smaller for contents with many disablers (alternatives), whereas alternatives (disablers) had little effect on the knowledge parameters for MP and MT (for AC and DA).

In Experiment 1, we contrast conditional rules "if p then q" and biconditional rules "if p then and only then q". As elaborated on below, we expect this manipulation to affect the τ parameters, but not the λ parameters. Experiment 1 thereby aims to validate that τ captures differences in reasoning form. For Experiment 2, we predict the converse pattern, via a manipulation of the expertise of the speaker of the conditional rule. We expect the manipulation to affect the λ parameter, but not the τ parameters. Experiment 1 also uses the four contents of Table 3, providing an opportunity to replicate the pattern predicted for the knowledge parameters just discussed.

These studies provide evidence for the validity of the different DSM parameters as measures of specific and separable cognitive processes. Based on them, we use the model to dissect a prominent phenomenon from the literature on conditionals, suppression effects (Byrne, 1989), into effects on knowledge-based processes versus form-based processes in Experiment 3, and we apply the model to a reanalysis of studies by Markovits, Brisson, and de Chantal (2015). Next, we compare the performance of the DSM and a number of Bayesian competitor models, such as Oaksford et al.'s (2000) model, in a goodness-of-fit meta-analysis. Finally, we compare the DSM to dual-process theories (Evans & Stanovich, 2013) and show that many criticisms of the latter do not apply to the DSM.

General Method

In all experiments, participants first worked on reduced inferences in what we call the knowledge phase and then on full inferences in what we call the rule phase. The two phases are separated by at least one week to avoid trivial carry-over effects. The reduced inferences consist of a minor premise and a conclusion; the probability of the conclusion has to be assessed. For example:

> Observation: A balloon is pricked with a needle.
> How likely is it that it will pop?

In the rule phase, participants work on full conditional inferences with conditional, a minor premise, and a conclusion, for example:

> Rule: If a balloon is pricked with a needle then it will pop.
> Observation: A balloon is pricked with a needle.
> How likely is it that it will pop?

In both phases, responses are given on a probability scale from 0% to 100%. Responses are divided by 100 for the analyses.

In each phase, participants provide responses for all four inferences per content. To obtain more reliable estimates, participants were also asked to provide responses for the so-called converse inferences MP', MT', AC', and DA' (Oaksford et al., 2000) in which the conclusion is presented in negated form. Continuing the above example, MP' for the full inference reads as follows:

> Rule: If a balloon is pricked with a needle then it will pop.
> Observation: A balloon is pricked with a needle.
> How likely is it that it will NOT pop?

For each content, the two responses to an inference and its converse were combined into one estimate.[3]

Each phase was split into two blocks. In each block, participants responded to the four inferences for each content in either original or converse form. For each inference and content, one member of each pair of inference and converse was randomly selected (e.g., either MP or MP') for the first block, whereas the other member was shown in the second block. Within each block, the four inferences per content were blocked and presented in random order. Order of contents was also randomly determined anew for each block.

There were two control groups, the *knowledge control group* and the *rule control group* to assess possible biases due to the fixed sequence of knowledge phase followed by rule phase. The knowledge control group was administered the knowledge phase twice, separated by at least one week. The rule control group was administered the rule phase as first phase, and participants had to appear for a second phase with unrelated tasks or without tasks.

An analysis of the data from these control groups showed that the pattern of results reported for the experimental groups in within-participants comparisons was also obtained when analyzing the data from the control groups in between-participants comparisons based on their first-phase data. Furthermore, in the knowledge control group, we found that responding to the knowledge phase items a second time led to a small but reliable increase of .02 to .03 on the response scale divided by 100. This increase was however negligible compared to the increase found in the rule phases of the experimental groups ($\geq$ .10). Modeling the former (small) increase explicitly within the DSM led to virtually the same pattern of results as reported here. For these reasons and for reasons of space, we decided to relegate the analysis of the control conditions to the online appendix.

Almost all participants in our studies were University-of-Freiburg students. Individuals with education in formal logic were not permitted to participate.

The raw data as well as complete analysis scripts including model fitting routines are available in the supplemental materials (available at https://osf.io/zcdfq/). The supplemental materials furthermore contains a single manuscript referred to throughout the article as the online appendix.

Experiment 1: Selective Influence on the τ Parameters

The first experiment contrasted standard conditionals in "if p then q" form with biconditionals in "if p then and only then q" form. From a logical perspective all four inferences are valid with a biconditional, whereas only two of them are valid with a conditional form. We therefore predicted that this manipulation should affect the τ parameters, but there was no reason to expect an effect on the overall weight given to form-based information as quantified by the λ parameters. Using the contents of Table 3, we also expected to replicate the effects of disablers and alternatives already described above.

Method

Participants. A total of 105 persons (35 per group) took part in the first phase of the experiment. Of these, 86 (mean age 22.2 years, *SD* = 2.7) participated in all phases of the experiment; 31, 26, and 29, in order, in the experimental group, the knowledge control group, and the rule control group. After finishing all phases of the experiment, participants received a monetary gratification of 16€.

Materials. All materials were presented in German, participants' mother tongue. The conditionals in "if-then" form are the ones shown in Table 3. For the biconditionals, the form "if p then and only then q" was used, for example: "If a predator is hungry then and only then it will search for prey"; see online appendix for the complete list of materials.

Procedure. The procedure followed the "General Method" described above with the following changes. To manipulate form, two rule phases were established (i.e., there were three sessions in total), in the first rule phase two randomly selected contents were shown in the conditional form and the other two in the biconditional form. This mapping was reversed for the second rule phase. In total, members of the experimental group responded to 32 items in the knowledge phase, generated by crossing 4 contents, 4 inferences (MP, MT, AC, & DA), and 2 conclusion polarities (original vs. converse), and to 32 items per rule type (conditional vs. biconditional) spread across two rule phases. After aggregating across original and converse inference, $3 \times 16 = 48$ data points remained per participant.

Members of the knowledge control group were administered the knowledge phase in all three sessions. Members of the rule control group started with the rule phases in the first and second session, which were otherwise identical to the two rule phases of the experimental group, but they also had to appear for a third session.

Results

Observed Data. The observed data are displayed in the upper row of Figure 2 for knowledge phase and the rule phases, separately for conditional and biconditional rules. The data were submitted to a repeated-measures ANOVA with factors type (without conditional, conditional inferences, biconditional inferences), inference (MP, MT, AC, DA), and content.

[3] E.g., for MP: (E(MP)+[1−E(MP')])/2). The two responses are perfectly consistent with each other if their sum, s, equals one (for responses divided by 100). Consistency, quantified by $(1 - |s - 1|)$, was used as a weight in the weighted least-squares estimation of the DSM parameters.

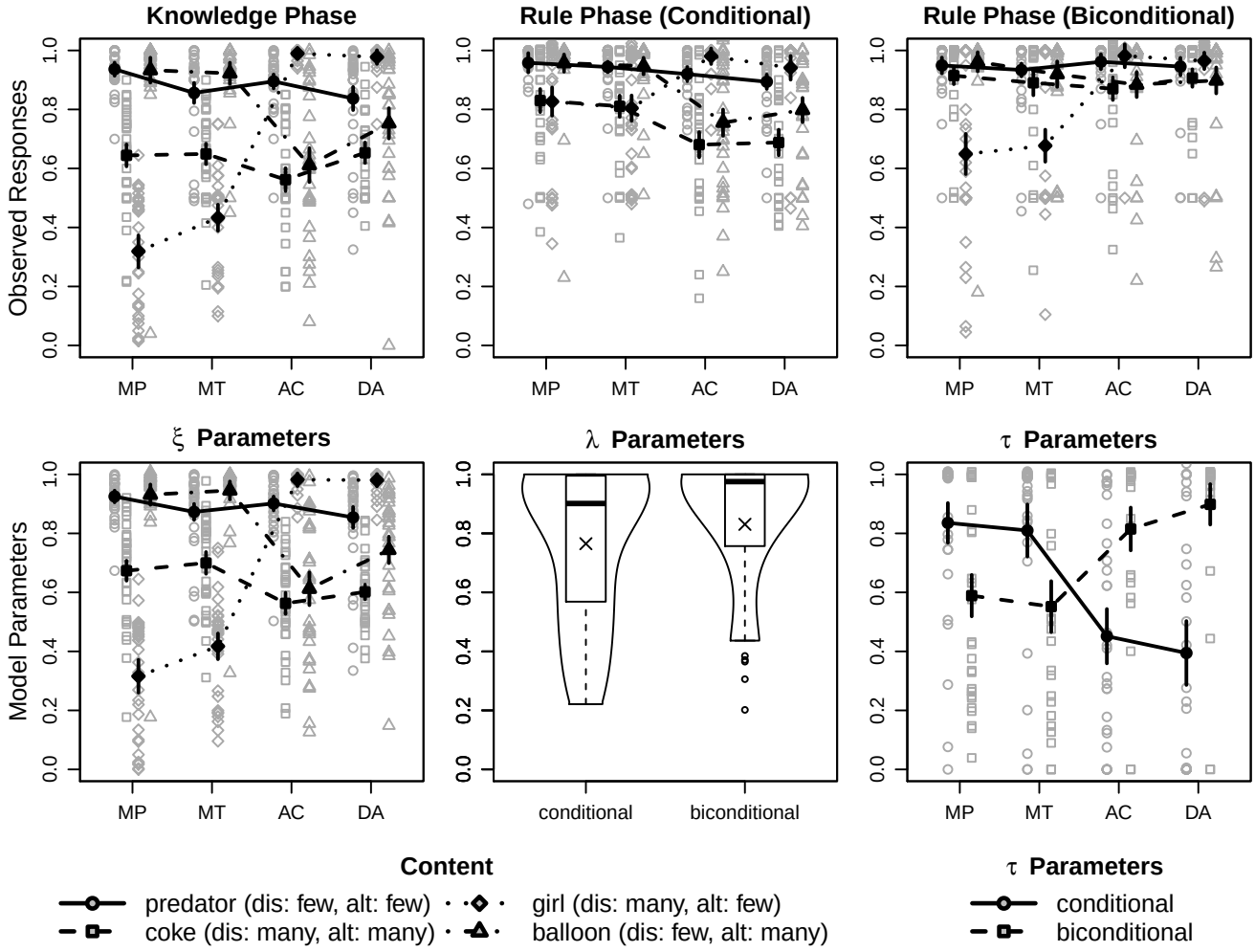

Figure 2. Observed responses (upper row) and model parameters (lower row) from Experiment 1. Mean values are displayed in black, error bars show per-plot difference-adjusted 95%-Cousineau-Morey-Baguley intervals (Baguley, 2012). Individual participants' values are displayed as in Figure 1. The λ parameters (lower row, middle panel) are displayed in a violin plot (Hintze & Nelson, 1998) in which the outlines show the density, the boxplots the first, second (i.e., median), and third quartiles, and the × the mean. dis = disablers; alt = alternatives.

There was a significant main effect of type, $F(1.68, 50.28) = 61.74$, $\eta_G^2 = .12$, $p < .0001$,[4] and a significant interaction of type and inference, $F(3.70, 111.04) = 12.08$, $\eta_G^2 = .03$, $p < .0001$: Mean endorsement was larger by .13 for inferences with conditional rule than for reduced inferences, $t(60) = 10.80$, $p_H < .0001$, and this increase, Δ, was more pronounced for MP and MT than for AC and DA, $\Delta_{\text{MP and MT}} = .16$, $\Delta_{\text{AC and DA}} = .09$, $t(180) = 4.53$, $p_H < .0001$. Both increases were individually significant (both $p_H < .0001$). Switching from conditional to biconditional rules led to a further increase in overall endorsement by an amount of .04, $t(60) = 2.61$, $p_H = .01$.

As can be seen in Figure 2, the expected effects of disablers and alternatives were also present across all inference

[4]Throughout the manuscript degrees of freedom (*df*) are Greenhouse-Geisser corrected in ANOVAs for repeated-measures factors with more than two levels. Furthermore, we report η_G^2 as effect size measure for ANOVAs following Bakeman's (2005) recommendations. Contrasts are obtained using the methods implemented in `lsmeans` (Lenth, 2015). For ANOVAs, contrasts use uncorrected Satterthwaite approximated *df*. To control for multiple testing, *p*-values of follow-up contrasts are corrected using either the Bonferroni-Holm correction (Holm, 1979), denoted p_H, or, in case all follow-up tests share the same *df* and do not involve one-tailed tests, a generalization of the Bonferroni-Holm method which takes the correlation among parameters into account (method `free` in package `multcomp`, Bretz, Hothorn, & Westfall, 2011), denoted p_F. Tests are one-tailed throughout the manuscript for the following hypotheses: (a) endorsements increase from knowledge phase to rule phase, (b) in particular, they increase for MP and MT, and

types and most pronounced in the knowledge phase. Contents with many disablers (i.e., the coke and girl content) were associated with depressed endorsement for MP and MT, whereas contents with many alternatives (i.e., the coke and balloon content) were asssociated with depressed endorsement for AC and DA, leading to a substantial interaction of inference and content, $F(4.11, 123.23) = 62.22$, $\eta^2_G = .26$, $p < .0001$.

In addition to the effects reported above, all other effects of the ANOVA were significant, including the three-way interaction of type, inference, and content, $F(8.50, 255.02) = 15.53$, $\eta^2_G = .08$, $p < .0001$. The just-reported effects were thus further modified by a higher-order interaction. Note, however, that the reported patterns hold quite consistently across conditions.

Dual-Source Model. We fitted the DSM to the 48 data points of each individual participant using 20 parameters: 4 $\times$ 3 parameters underlying ξ plus 2 $\times$ 4 $\lambda\tau$ parameters from which we obtained two sets of one λ and four τ parameters each, one set for the conditional rule and one set for the biconditional rule. Overall model fit was good, mean $R^2 = .90$ ($SD = .06$), mean and individual parameter estimates are shown in the lower row of Figure 2.

Results confirmed the prediction of selective influence. The effect of rule type (conditional vs. biconditional) on the mixture weight parameters λ was small and not significant in a repeated-measures ANOVA with factor rule type, $F(1, 30) = 3.15$, $\eta^2_G = .02$, $p = .09$. In contrast, a repeated-measures ANOVA of the τ parameters with factors rule type and inference revealed a small main effect of rule type, $F(1, 30) = 4.92$, $\eta^2_G = .02$, $p = .03$, moderated by a substantial interaction with inference, $F(2.10, 63.10) = 27.54$, $\eta^2_G = .20$, $p < .0001$. There was no main effect of inference, $F < 1$. Taken together, rule type did not affect the weighting of the two types of information; instead, it affected the form-based information.

Follow-up contrasts revealed that the τ parameters were lower for the biconditionals than for the conditionals by a mean amount of $-.25$, $t(83.78) = -4.51$, $p_H < .0001$, for MP and MT, whereas they were larger for AC and DA by a mean amount of .43, $t(83.78) = 7.76$, $p_H < .0001$. The τ parameters for the conditional inferences showed the pattern usually obtained for abstract conditionals (see lower right panel of Figure 2).

The knowledge parameters ξ mimicked the pattern of endorsements observed in the knowledge phase: A repeated-measures ANOVA of the ξ parameters with factors inference and content revealed a significant interaction, $F(4.41, 132.30) = 119.61$, $\eta^2_G = .56$, $p < .0001$. ξ parameters of MP and MT were smaller for contents with many disablers than for contents with few disablers by a mean amount of $-.39$, $t(161) = -19.76$, $p_F < .0001$, whereas there was no such difference for AC and DA, the mean difference being .00, $t(161) = 0.19$, $p_F = .85$. The ξ parameters of AC and DA were smaller for contents with many alternatives than for contents with few alternatives by a mean amount of $-.30$, $t(161) = -15.08$, $p_F < .0001$, and the ξ parameters of MP and MT were larger by a mean amount of .18, $t(161) = 9.06$, $p_F < .0001$.

Discussion

We found the predicted selective influence of the inference form manipulation on the parameters of the DSM. Contrasting conditional and biconditional inferences affected only the form-parameters τ, but not the mixture weights λ. These results are consistent with the DSM assumption that the role of a (bi)conditional rule in probabilistic (bi)conditional inferences is to provide content-independent form-based information that reasoners integrate with their background knowledge about the content of the inference.

That τ parameters of MP and MT were found lower for the biconditional than for the conditional was somewhat unexpected. This pattern does not conform to the usual interpretation of biconditionals according to which the biconditional rule licenses all four inferences alike. However, instead of the "canonical" form of biconditionals ("if and only if p then q"), we used a form that sounded more natural in a probabilistic setting ("if p then and only then q"). This form may have specifically emphasized the necessity of p for q on which AC and DA are based, while being more neutral towards the sufficiency of p for q on which MP and MT are based.

Experiment 2: Selective Influence on the λ Parameters

The second experiment sought evidence for the assumption that the λ parameter reflects the relative weight given to the form-based versus knowledge-based information. To this end, we employed a paradigm introduced by Stevenson and Over (2001) in which we manipulated – within-participants – the expertise with which a conditional was uttered: Some conditionals were uttered by an expert on the rule's content, others by a non-expert. For example, the conditional "If Anne eats a lot of parsley then the level of iron in her blood will increase" could have been stated by either a nutrition scientist or a drugstore clerk. As this manipulation does not change the form of the inferences, it should not affect τ. Rather, if a conditional is uttered by a non-expert (e.g., a drugstore clerk), reasoners should discount the rule and rely more strongly on their background knowledge of the subject matter than for rules stemming from an expert source. In terms of the DSM, this should lead to an effect on the λ pa-

(c) for AC and DA, (d) replicating the previous work, the increases in MP and MT are larger than those for AC and DA, (e) the effects of speaker expertise in Experiment 2. All other tests are two-tailed.

rameter for the relative weight given to form-based versus knowledge-based information.

A secondary goal was to see whether the applicability of the DSM would generalize to contents without strong causal links between antecedent and consequent.[5] Note that we nevertheless did not use abstract materials, but everyday contents for which participants prior knowledge was assumed to be rather vague. This ensured that the manipulation of speaker expertise could overshadow participants' prior knowledge. To develop appropriate materials we first performed a large online pilot study (N = 435) in which we tested 20 different conditionals constructed after Stevenson and Over (2001) for which participants were expected to have little prior causal knowledge. Results replicated the effect of speaker expertise reported by Stevenson and Over (2001) on MP such that endorsement was lower for conditionals uttered by non-experts compared to experts. However, we could not replicate such an effect on MT. Extending Stevenson and Over's work, we also tested AC and DA and found a similar effect on DA, but not on AC. For the current experiment, we selected the seven conditionals with particularly strong effects of speaker expertise. A full description of the pilot study can be found in the online appendix.

Method

Participants. Because the effects of speaker expertise observed in the pilot study were comparatively small, we collected more participants per group than in our other experiments. A total of 153 participants (51 per group) were sampled. Of these, 138 (mean age 21.9 years, SD = 3.4) participated in both phases of the experiment; 47, 46, and 45, in order, in the experimental group, the knowledge control group, and the rule control group. Participants received 12€ as monetary gratification.

Materials and Procedure. The procedure followed the "General Method" with a few changes. Six contents (out of the pool of seven) were randomly selected for each participant. In the knowledge phase, participants only saw the minor premise and were asked to rate the probability of a conclusion. The minor premise was labeled "situation"; for example,

> Situation: Anne eats a lot of parsley.
> How likely do you think it is that the level of iron in her blood will increase?

Participants provided 48 responses, obtained by crossing 6 contents, 4 inferences (MP, MT, AC, DA), and 2 conclusion polarities (original vs. converse).

In the rule phase, the same six contents were again used. For three randomly selected contents, the conditional rules were uttered by an expert (e.g., "A nutrition scientist says: If Anne eats a lot of parsley then the level of iron in her blood will increase."), and by a non-expert for the remaining three rules (e.g., "A drugstore clerk says: If Anne . . . "). The manipulation of speaker expertise was performed within-participants; the expertise assigned to each rule was randomized across participants. Note that speaker expertise was manipulated only in the rule phase and not in the knowledge phase. After aggregating across original and converse inference, participants provided 2 × 24 = 48 data points across knowledge and rule phase.

Participants in the knowledge control group responded to reduced problems in both sessions. Participants in the rule control group started with the rule phase in the first session which was identical to the rule phase of the experimental group, but they also had to appear for a second session.

Results

Observed Data. We submitted the observed data to a linear mixed model (LMM) with fixed effects for inference, phase (knowledge vs. rule), expertise (expert vs. non-expert), and their interactions. Note that for this analysis, contents that were assigned to experts or non-experts in the rule phase were assigned the same expertise status in the knowledge phase although we did not manipulate expertise in the knowledge phase (i.e., an effect of expertise in the knowledge phase was a priori impossible). Furthermore, we estimated crossed random effects for participant and content with maximal random slopes (i.e., random slopes for inference, expertise, phase, and their interactions for both random effects; Barr, Levy, Scheepers, & Tily, 2013; see online appendix for more details). Each participant provided three responses per cell of the design. The estimated marginal means of the LMM are displayed in Figure 3, upper row.

The only significant effect to emerge was a main effect of phase, $\chi^2(1) = 13.35$, $p = .0002$, all other $p > .19$. Responses were larger by a mean amount of .12 for full inferences in the rule phase compared to reduced inferences in the knowledge phase.

We fitted a separate LMM to the rule-phase data in which expertise had been manipulated. This provided some evidence for an effect of expertise, $\chi^2(1) = 3.66$, $p = .06$; $p > .49$ for all other effects and interactions. Follow-up contrasts showed that only for MP was there evidence for an effect of speaker expertise (difference = .05), $z = 2.41$, $p_H = .03$, but not for the other three inferences, all $p_H > .23$.

Dual-Source Model. The DSM for the 48 data points per participant uses 26 parameters: 6 × 3 parameters underlying ξ plus 2 × 4 $\lambda\tau$ parameters from which we obtained

[5]This goal was also reflected in the statistical analysis: Instead of treating content as a fixed effect, we now treated it as a random effect to allow us to generalize conclusions to the population of similar contents. More specifically, we treated both participants and contents as *crossed random effects* (Baayen, Davidson, & Bates, 2008). For examples in the reasoning domain see Haigh, Stewart, and Connell (2013) and Singmann, Klauer, and Over (2014).

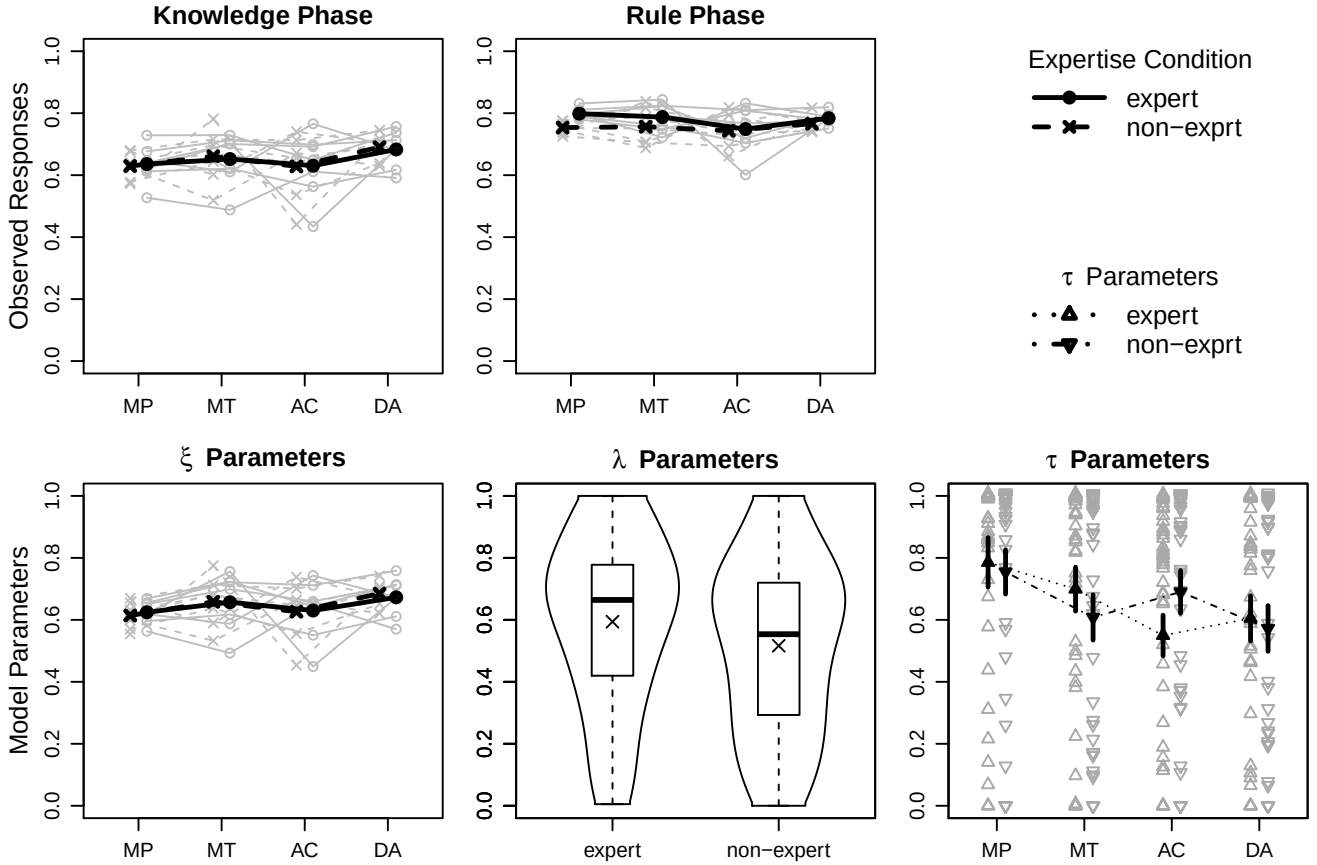

Figure 3. Observed responses (upper row) and DSM parameters (lower row) for Experiment 2. Estimated marginal means from LMMs are displayed in black, grey lines show LMM estimates for the random effect of content. The λ and τ parameters are displayed as in Figure 2.

two sets of one λ and four τ parameters, one set for each expertise condition. Overall model fit was acceptable, mean $R^2 = .82$ ($SD = .13$), mean and individual parameter estimates are shown in Figure 3, lower row.

A repeated-measures ANOVA of the λ parameters with factor expertise showed that mean λ was significantly larger for experts than for non-experts, $F(1, 46) = 6.88$, $\eta^2_G = .02$, $p = .01$. In contrast, a repeated-measures ANOVA of the τ parameters with factors expertise and inference revealed only a main effect of inference, $F(2.55, 117.45) = 5.08$, $\eta^2_G = .03$, $p = .004$, but no main effect of expertise, $F < 1$, nor an interaction of both factors $F(2.76, 127.10) = 1.93$, $\eta^2_G = .01$, $p = .13$. Considering the main effect of inference, the pattern of the τ parameters descriptively followed the pattern usually obtained for abstract contents (see lower right panel of Figure 3). Taken together, the manipulation of speaker expertise affected only the weighting of the two types of information (λ), but not the form-based information (τ).

Discussion

We found the predicted selective influence of speaker expertise on the parameters of the DSM. The expertise with which the conditional was uttered affected only the weight parameter λ, but not the form parameters τ. Experiment 2 also replicated Stevenson and Over's (2001) classical expertise effect for MP. To our knowledge, this is the first replication of the effect. Going beyond replication, the present results suggest that the effect is driven by the mixture weight: Some participants discount a low-expertise conditional, relying on their background knowledge more strongly for such conditionals than for high-expertise conditionals.

In addition, Experiment 2 provided evidence for the generalizability of the DSM. In contrast to previous studies, we selected contents from a wider set and treated them as random effects in the analysis. Furthermore, we used materials that differed from the ones typically employed in previous experiments on the DSM insofar as we used materials for which participants were expected to have little prior causal knowledge. In consequence, the responses to the different inferences were all more or less on the same level (see Figure 3). Importantly, the form parameters τ still tended to exhibit the pattern usually obtained for abstract conditional inferences. This finding provides further support for the interpretation of the τ parameters as reflecting the impact of

content-independent information based on the form of an inference.

The DSM in its current form does not permit an influence of the conditional rule on the participants' background knowledge. It is possible that a Bayesian model of the effects of speaker expertise could be developed in which updating by a conditional rule changes the knowledge base via an updating process that leads to different changes in one's knowledge base as a function of whether the conditional is uttered by an expert or by a non-expert. As discussed in the introduction, there is, however, currently no agreed-upon or normative way of Bayesian updating by a conditional premise (see below for three descriptive approaches).

Together with Klauer et al.'s results (2010), the results presented so far provide evidence for the theoretical assumptions underlying the DSM and for the intended psychological interpretations of the model parameters. At least two dissociable and quantifiable cognitive processes contribute to probabilistic conditional reasoning, plus an independent and dissociable weighting process determining how the two types of information are integrated. In the next experiment, we move from validation to applying the DSM as a measurement model to dissect the classical suppression effects first described by Byrne (1989) into possible effects on the knowledge-based component, the form-based component, or their mixture.

Experiment 3: Dissecting Suppression Effects

Human reasoning is *defeasible* or *non-monotonic*: New premises can render a previously acceptable inference unacceptable. A classical demonstration of non-monotonicity goes back to Byrne (1989) who compared acceptance rates of conditional inferences with and without an additional premise. For example, given Premises 1 and 3 below, most reasoners accept the Conclusion 4 in an MP inference. Upon adding Premise 2, the acceptance of the conclusion usually drops dramatically.

1. If a balloon is pricked with a needle then it will quickly lose air.

2. If a balloon is inflated to begin with then it will quickly lose air.

3. A balloon is pricked with a needle.

4. The balloon quickly loses air.

MP and MT require that p is sufficient to bring about q, and sufficiency is questioned by disablers. AC and DA require that p is necessary for q to occur, and necessity is questioned by alternatives. Premises such as Premise 2 suggesting a disabling condition (the balloon is not inflated in the first place) reduce acceptance of the MP and MT inferences with little effect on AC and DA. Conversely, adding a premise suggesting an alternative cause (e.g., Premise 2': If a balloon is inflated too much and bursts then it will quickly lose air) decreases acceptance of AC and DA with little effect on MP and MT (Byrne, Espino, & Santamaria, 1999; Chan & Chua, 1994; Neth & Beller, 1999).

Several explanations have been offered. One possibility retains monotonicity of human reasoning and claims that the two rules are combined into one such as "If the balloon is pricked with a needle and it is inflated to begin with then it will quickly lose air" (e.g., Byrne, 1989; Stenning & van Lambalgen, 2010) or similarly that they are combined in one mental model (Johnson-Laird & Byrne, 2002), changing the formal or semantic structure of the argument. Another possibility, based on the pragmatics of natural language, is that the second premise undermines the belief in the truth of the conditional rule (Bonnefon & Politzer, 2010), undermining all inferences that might be drawn from it formally. A third possibility exemplified by Oaksford and Chater (2007, chap. 5) is that the second premise alters the knowledge base on which the assessment of the probability of the conclusion rests.

In terms of the DSM, the first possibility should lead to a change in the profile of τ parameters across inferences, the second to a decrease in the weight λ given to the form-based information, the third to an effect on the knowledge parameters ξ. Note that the three possibilities by no means exclude each other. Experiment 3 aims at assessing the extent to which suppression effects reflect effects on the form-based information (τ), on the weight given to such information (λ), and/or on the knowledge parameters ξ in probabilistic reasoning.

A baseline group followed the procedures described in the section "General Method" above; members of two additional groups, the disablers group and the alternatives group, worked on the same problems with additional information presented for each problem. In the disablers group, the additional information specified disablers; in the alternatives group, alternative causes. The additional information was always presented in both knowledge and rule phase to enable us to assess the influence of the additional information on the full set of DSM parameters.

Experiment 3 comprised two independent replications, Experiment 3a and 3b. Experiment 3a was performed first, Experiment 3b was performed second with the goal of replicating the findings and of ruling out some alternative explanations. To this end, we validated the additional information used in Experiment 3a prior to running Experiment 3b in a separate study. Experiment 3b also implemented a separate rule control group and a separate knowledge control group for each of the three experimental groups (see online appendix). As the pattern of results obtained in the two ex-

periments was virtually indistinguishable, we pooled the data sets and only report the pooled analysis here, retaining experiment as a factor in the analyses (see the online appendix for the separate analyses).

Method

Participants. In Experiment 3a, a total of 80 persons participated in the first phase of the experiment; 26, 27, and 27, in order, in the baseline group, disablers group, and alternatives group. Of these, 77 (mean age 23.5 years, *SD* = 4.5) participated in both phases of the experiment (3 participants in the alternatives group did not appear for the second session).

In Experiment 3b, a total of 281 persons participated in the first phase; in the experimental groups, in order, 31, 31, and 31 in the baseline group, disabler group, and alternatives group; in the knowledge control groups, 32, 31, and 31; in the rule control groups, 31, 32, and 31. Of those, 273 (mean age 21.5 years, *SD* = 3.2) participated in both phases of the experiment (29, 31, and 31 in the experimental groups; 31, 31, and 30 in the knowledge control groups; 29, 32, and 29 in the rule control groups). In both experiments participants received 12€ for their participation.

Materials and Procedure. We used the four contents displayed in Table 3 (we slightly adapted contents 2 and 3 such that we were able to present both disablers and alternatives). Pilot work pretested different ways of presenting the additional information. Presenting the additional information as additional conditionals as done by Byrne (1989) did not lead to suppression effects of similar size as the ones reported by Byrne; the obtained effects were small or absent.[6] Therefore, we decided to present three additional counterexamples in a form clearly signaling their causal influence on the conditional, using "only … if" for disablers and "also … if" for alternative causes. For example, MP for the coke content in the disablers group was presented in the following way in Experiment 3b:

> *Rule:* If a person drinks a lot of coke then the person will gain weight.
> *Observation:* A person drinks a lot of coke.
> How likely is it that the person will gain weight?
> Please note:
> A person *only* gains weight if
>
> - the metabolism of the person permits it,
> - the person does not exercise as a compensation,
> - the person does not only drink diet coke.

In the alternatives group, this content was presented in the following way:

> *Rule:* If a person drinks a lot of coke then the person will gain weight.
> *Observation:* A person drinks a lot of coke.
> How likely is it that the person will gain weight?
> Please note:
> A person *also* gains weight if
>
> - the person eats a lot,
> - the person has metabolic problems,
> - the person hardly exercises.

Prior to Experiment 3b we assessed the validity of the presented additional information in a pretest. The full list of the items used in Experiments 3a and 3b (including the additional information and a description of the pretest) can be found in the online appendix.

The procedure otherwise followed the "General Method" with the only difference that the disablers and alternatives were presented alongside the inferences in the disablers and alternatives groups, respectively, as exemplified above in both knowledge and rule phases.

Participants responded to 32 items in each phase obtained by crossing 4 contents, 4 inferences, and 2 conclusion polarities. After aggregating original and converse inference, there were $2 \times 16 = 32$ data points per participant.

Results

Observed data. The upper rows of Figure 4 present mean endorsements as a function of group and inference (see online appendix for plots further split up by contents). A mixed-effects ANOVA with within-participants factors phase (knowledge vs. rule), inference (MP, MT, AC, and DA), and content (see Table 3), and between-participants factors group (baseline, disablers, and alternatives), and experiment (Experiment 3a vs. 3b), was run to see whether suppression effects were observed in the rule phase and perhaps already in the knowledge phase. Note that for this ANOVA and all further ANOVAs reported for Experiment 3, the effect of experiment did not reach significance, all $F < 1.3$, all $p > .08$.

The signature of suppression effects is that disablers suppress MP and MT, but not AC and DA, and vice versa for alternatives. Consistent with this signature, we found an interaction of inference and group, $F(3.69, 298.65) = 40.51$, $\eta^2_G = .07$, $p < .0001$, qualified by a significant three-way interaction with phase, $F(4.91, 397.72) = 2.99$, $\eta^2_G = .002$, $p = .01$. This latter interaction indicated that suppression effects differ across phases.

[6] Asking participants to assume the truth of the premises and to judge the truth of the conclusion as done by Byrne (1989), we replicated her results.

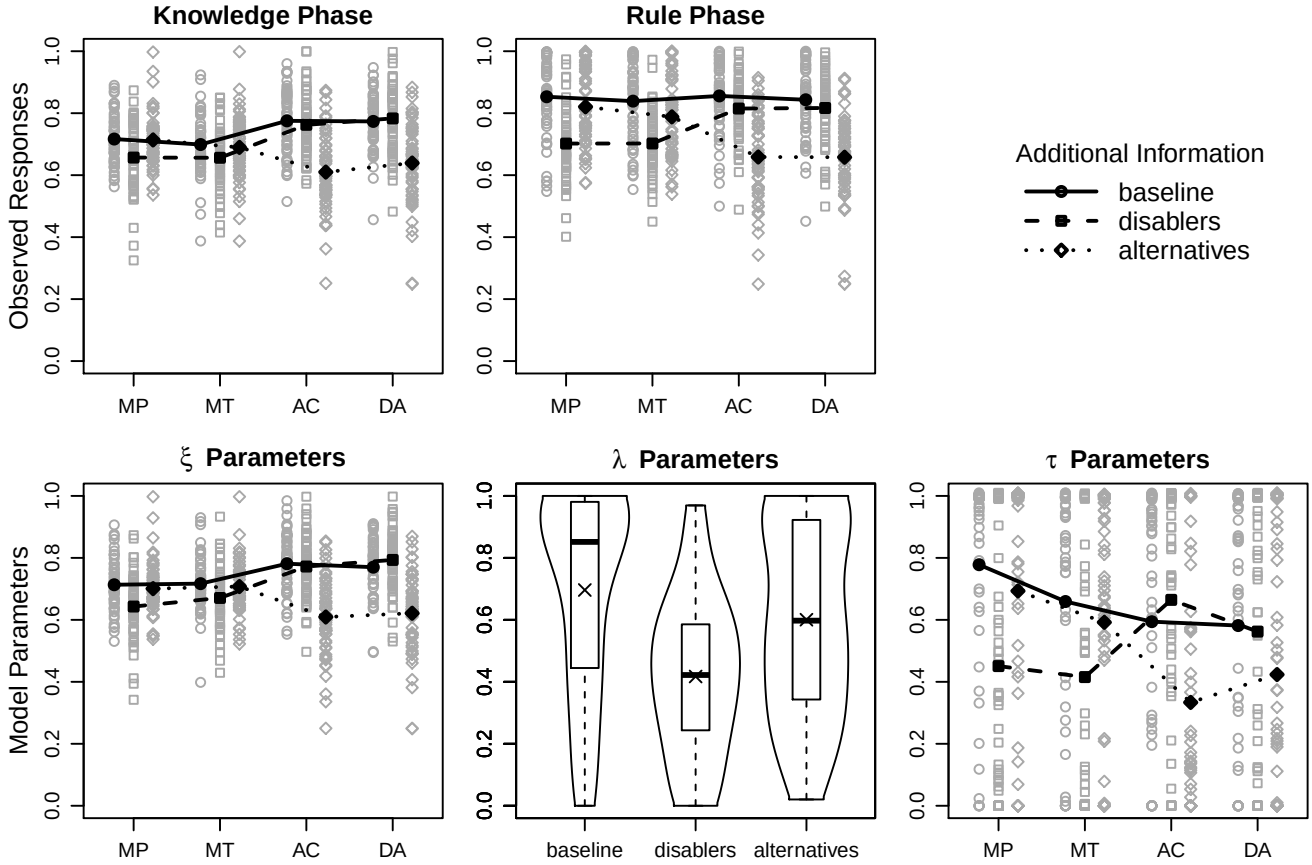

Figure 4. Observed responses (upper row) and model parameters (lower row) from Experiments 3a and 3b combined. Mean and individual participants' data points are shown as in Figure 2.

As can be seen in Figure 4 (upper row), endorsement of MP and MT was indeed decreased in the disablers group relative to the baseline group, although this suppression effect was smaller in the knowledge phase, $M = -.05$, $t(412.69) = -2.87$, $p_H = .03$, than in the rule phase, $M = -.14$, $t(412.69) = -8.14$, $p_H < .0001$ (p_H of difference = .0001). Analogously, endorsement of AC and DA was decreased in the alternatives group relative to the baseline group, and this suppression effect was of similar size in knowledge phase, $M = -.15$, and rule phase, $M = -.19$, both $t < -8.1$, both $p_H < .0001$ (p_H of difference = .07). In contrast, in neither phase did disablers suppress AC and DA or alternatives MP and MT, all $|t| < 2.26$, all $p_H > .09$. Taken together, there were clear suppression effects showing the classical signature.

Dual-source model. We fitted a DSM to each participant's 32 data point using 16 parameters: 4×3 parameters underlying ξ, and 4 $\lambda\tau$ parameters from which we obtained one λ and four τ parameters. The overall model fit was good, mean $R^2 = .89$ (SD = .10). Parameter estimates are displayed in the lower row of Figure 4 as a function of inference and group.

As can be seen in the figure, disablers and alternatives suppress the τ parameters of the "attacked" inferences, that is of MP and MT for disablers, and of AC and DA for alternatives. Disablers also undermine the overall weight λ for the rule and all inferences based on it, whereas alternatives also suppress the knowledge parameters ξ for AC and DA.

Knowledge-parameters ξ**.** A mixed-effects ANOVA with between-participants factors group and experiment, and within-participants factors content and inference found a significant interaction of group and inference, $F(3.61, 292.75) = 32.96$, $\eta^2_G = .08$, $p < .0001$. According to follow-up contrasts, alternatives reduced ξ parameters for AC and DA by $M = -.16$ relative to the baseline group, $t(277.78) = -9.99$, $p_H < .0001$, but they left MP and MT unaffected, $M = -.01$, $t(277.78) = -0.59$, $p_H > .99$. Disablers exerted a small suppression effect on MP and MT, $M = -.06$, $t(277.28) = -3.72$, $p_H = .0007$ that was significantly smaller than the just-reported suppression effect of alternatives on AC and DA, $t(489.14) = 5.21$, $p_H < .0001$. Disablers left AC and DA unaffected, $M = .01$, $t(277.78) = 0.40$, $p_H > .99$.

Mixture-weight λ. An ANOVA with factors group and experiment showed only a significant main effect of group, $F(2, 162) = 12.51$, $\eta^2_G = .13$, $p < .0001$. The λ parameters were smallest in the disablers group, where they were significantly smaller than in the baseline group and in the alternatives group by mean amounts of $-.28$, $t(162) = -4.90$, $p_F < .0001$, and $-.19$, $t(162) = -3.25$, $p_F = .003$, respectively. The λ parameters of the latter two groups did not differ significantly from each other, $t(162) = 1.61$, $p_F = .11$.

Form parameters τ. A mixed-effects ANOVA with between-participants factors group and experiment, and within-participants factor inference showed a significant interaction of group and inference, $F(5.07, 410.28) = 7.06$, $\eta^2_G = .06$, $p < .0001$. Like for the ξ parameters, disablers suppressed the parameter values of MP and MT by $M = -.28$ relative to the baseline group, $t(463.37) = -5.23$, $p_H < .0001$, but they did not affect AC and DA, $M = .03$, $t(463.37) = 0.52$, $p_H = .73$. Analogously, alternatives suppressed AC and DA compared to the baseline by an amount of $M = -.21$, $t(463.37) = -3.89$, $p_H = .0005$, but did not affect MP and MT, $M = -.07$, $t(463.37) = -1.33$, $p_H = .55$. Unlike for the ξ parameters, the size of the suppression effect of disablers on MP and MT did not differ from that of the alternatives on AC and DA, $t(647.17) = -0.91$, $p_H = .73$. In addition, τ parameters in the baseline group again exhibited the pattern usually obtained for abstract contents.

Discussion

Experiment 3 considered suppression effects (Byrne, 1989) in the new paradigm using reduced and full inferences. Disablers suppressed endorsement of MP and MT primarily for the full inferences, whereas alternatives suppressed AC and DA equally strongly for reduced and full inferences.

The DSM dissects the suppression effects into three components: First, disablers and alternatives suppress the form-based evidence brought about by the "attacked" inferences (i.e., MP and MT for disablers and AC and DA for alternatives). This is in line with traditional accounts of the suppression effect (Byrne, 1989; Johnson-Laird & Byrne, 2002; Stenning & van Lambalgen, 2010). For example, suppression effects can be explained by assuming that disablers d are integrated into the rule, "if p then q" so that reasoners proceed from the rule "if $p \wedge \neg d$ then q" whereas alternative causes a are integrated so that reasoners proceed from the rule "if $p \vee a$ then q". These conditional rules, even if interpreted biconditionally (as "if-and-only-if then" rules), no longer license the attacked inferences, but still warrant the other ones.

Second, alternatives and disablers decreased the knowledge-based support of the attacked inferences, the effect of alternatives being much larger than that of disablers. This is in line with probabilistic accounts (Oaksford & Chater, 2007), but also with recent findings suggesting that reasoners neglect alternative causes, but not disablers, in probabilistic inferences with causal conditionals (Fernbach & Erb, 2013). In consequence, stating alternatives explicitly provides new, up to now unheeded information that prompts updating. This accounts for the observed effects on the knowledge parameters relative to the baseline group. In contrast, disablers are to some extent taken into account spontaneously and therefore trigger little updating relative to the baseline group.

Third, disablers also reduce the overall weight λ of the rule and all formal inferences based on it. Note that disablers directly discredit the conditional relationship between p and q as stated in the rule, whereas alternatives have no immediate relevance for the rule and its validity. This effect of disablers may thus resemble the effect of speaker expertise considered in Experiment 2.

The present analysis is the first to suggest that suppression effects can be decomposed into separate contributions of different kinds of processes. Some reflect effects on the knowledge base on which endorsements are based. Such effects can be directly observed in the endorsements of the reduced inferences. Additional processes related to the rule and the form of the full inferences come into play to shape the pattern of inferences in the rule phase with full inferences. The former effects are completely consistent with extant Bayesian accounts of suppression effects; the effects on the τ parameters are in line with traditional accounts of them (Byrne, 1989; Johnson-Laird & Byrne, 2002; Stenning & van Lambalgen, 2010); the effects on the overall weight λ given to the form-based information suggest a link between the speaker-expertise effect and the classical suppression effect.

Applying the DSM to a Reanalysis of Markovits et al. (2015)

The DSM can also be used outside our experimental paradigm to provide a unifying theoretical perspective. In a recent study, Markovits et al. (2015) compared probabilistic reasoning and deductive reasoning in which participants are asked to assume the truth of the premises. The studies used two different fictitious contents involving an alien planet for which participants had no prior knowledge. For each of the two different contents, participants first responded to an AC problem in Experiment 1 or to a DA problem in Experiment 2. They then received relevant frequency information about the occurrence of pq and $\neg pq$ cases in Experiment 1 or about the occurrence of $\neg pq$ and $\neg p \neg q$ cases in Experiment 2. Finally, they were asked to assess the same inference they had previously responded to a second time.

The frequency information for the two contents contrasted a high-probability condition with high $P(p|q)$ (Exp. 1) or high $P(\neg q|\neg p)$ (Exp. 2) and a low-probability condition. Markovits et al. predicted that under deductive instructions, acceptance of the initial inference should decrease in both

Table 4
Observed data from Markovits et al. (2015) and corresponding DSM predictions.

	Deductive Condition				Probabilistic Condition			
	First Response		Second Response		First Response		Second Response	
	High[a]	Low[b]	High[a]	Low[b]	High	Low	High	Low
				— Observed data —				
AC	.53	.55	.27	.09	.70	.71	.74	.34
DA	.56	.56	.16	.21	.59	.65	.64	.33
				– DSM predictions —				
AC	.57	.57	.30	.11	.68	.68	.58	.44
DA	.53	.53	.30	.10	.65	.65	.56	.41

Note. AC data from Study 1; DA data from Study 2. The first response was given before frequency information was provided, the second response was given after the frequency information was provided. The DSM used 7 parameters to fit the 16 observed data points with R^2 = .86.
[a]High-probability condition. [b]Low-probability condition.

conditions because both suggested the existence of counterexamples. In contrast, under probabilistic instructions, the initial estimates should decrease only in the low-probability condition, but not in the high-probability condition.

The results are reproduced in the upper part of Table 4 and showed the expected dissociation of probabilistic and deductive reasoning as a function of probability condition: Under deductive instructions, acceptance rates decreased from first to second response irrespective of probability condition. In contrast, for probabilistic instructions, endorsement decreased only for the low-probability condition, but not for the high-probability condition.

In modeling these data, we assume that the first responses without frequency information reflect only the form-based component of the DSM (see also Singmann, Klauer, & Over, 2014). In contrast, given frequency information, the second responses should again reflect a mixture of the form-based and knowledge-based components. Remember furthermore that when uncertain about the validity of an inference x (with probability $1 - \tau(x)$), reasoners are assumed to fall back on their background knowledge in probabilistic assessment of the conclusion. Under deductive instructions, it makes more sense to assume that the inference is simply not drawn when it is not clearly perceived as valid. Consequently, while the full DSM (Equation 6) should hold for responses in the probabilistic condition, the $(1 - \tau(x))$ cases are removed for modeling responses from the deductive condition:

$$\text{Response}(C, x) = \lambda \times \tau(x) + (1 - \lambda) \times \xi(C, x), \qquad (9)$$

The resulting DSM uses seven free parameters to fit the 16 data points shown in Table 4 (upper part): Two weight parameters λ, one for the deductive and one for the probabilistic condition, two form parameters τ, one for AC and one for DA, and three knowledge parameters ξ, one each for the high-probability condition, the low-probability condition, and for the first responses under probabilistic instructions (i.e., a "probabilistic baseline").

The DSM described the data reasonably well as shown by the predicted values given in the lower part of Table 4 and the overall goodness-of-fit index, R^2 = .86. Importantly, the estimates of the λ parameters were in line with our a priori expectations: λ was much larger in the deductive condition than in the probabilistic condition, λ = .78 and λ = .19, respectively. The τ estimates for AC and DA were similar: τ(AC) = .57 and τ(DA) = .53. The ξ parameter estimates were .24 for the high probability condition, .00 for the low-probability condition, and .25 for the "probabilistic baseline".

The reanalysis exemplifies that the dual-source framework can provide a unifying theoretical view on seemingly disparate theoretical positions. In particular, the two reasoning modes, probabilistic and deductive reasoning, need not necessarily be associated with different processes. In our model both types of reasoning use the same form-based and knowledge-based information with the major difference being their differential weighting in determining observed responses. Note, however, that the DSM agrees with the conclusion of Markovits et al. (2015) that deductive updating is not purely Bayesian.

Finally, consider a pattern in the second responses that may be responsible for the only satisfactory R^2 value: Under deductive instructions, acceptance of DA slightly increases in the low-frequency relative to the high-frequency condition; under probabilistic instructions, it decreases. The latter, pronounced decrease forces the DSM to also predict a decrease in second DA responses under deductive instructions. A further test of the DSM in this application would therefore be to replicate the DA condition to see if the pattern found by Markovits et al. replicates or if one also finds a decrease from high-probability to low-probability condition in second

Table 5
Overview of data sets used in the meta-analysis and number of free parameters of each model

Data set	n	Contents	Data Points	DSM	PROB / KL	EX-PROB
This Manuscript:						
Exp. 1 (conditional inferences)	31	4	32	16	16	20
Exp. 2	47	6	48	22	24	30
Exp. 3a (baseline condition)	26	4	32	16	16	20
Exp. 3b (baseline condition)	29	4	32	16	16	20
Klauer et al. (2010):						
Exp. 1 (phases 1 & 2)	15	4	32	16	16	20
Exp. 3 (conditional inferences)	18	5	40	19	20	25
Exp. 4 (conditional inferences)	13	4	32	16	16	20
Total n	179					

Note. DSM = dual-source model; PROB = probabilistic model of Oaksford et al. (2000); KL = Kullback-Leibler model; EX-PROB = extended probabilistic model of Oaksford and Chater (2007). See online appendix for more details.

responses under deductive instructions as the DSM predicts.

Goodness-of-Fit Meta-Analysis

Finally, we engaged in a model-comparison exercise (Heathcote et al., 2015) to assess the DSM relative to purely Bayesian approaches. We compare the DSM and Bayesian competitor models using data from both the current manuscript and from Klauer et al. (2010).

Data Sets and Competitor Models

We used the seven data sets (or parts of data sets) with 179 participants that implement the procedure described in section "General Method". More specifically, we excluded conditions that employed other types of conditionals (e.g., the biconditionals of Exp. 1) or that altered the presented inferences (e.g., in the additional information conditions of Exp. 3). While it was clear to us how such manipulations are to be modeled with the DSM, it was not clear how to model them by the competitor models (we do not doubt that they could be somehow modeled).

Bayesian models assume that observed endorsements reflect the reasoners' assessments of conditional probabilities of the conclusion given minor premise. This would be true of both the reduced inferences of the knowledge phase and the full inferences of the rule phase. Moving from knowledge phase to rule phase can be naturally modeled as updating the probability distribution driving one's assessments to accommodate the additional premise, the conditional rule. As already discussed, there is however no agreed-upon or normative mechanism of updating by a conditional premise.

Several descriptive models have been proposed for such updating in the Bayesian framework. A first model, termed PROB here, has already been discussed in the introduction. It assumes that for each content, introducing the conditional is modeled by a new exceptions parameter e' in Oaksford et al.'s (2000) model with $e' < e$, and thus e' and e are separate exceptions parameter for the knowledge-phase and rule-phase data, respectively.

An extension of PROB, termed EX-PROB here, was proposed by Oaksford and Chater (2007, pp. 126; see also Oaksford & Chater, 2013). This extension is motivated by the observation that MP is usually endorsed much more strongly than MT and employs two additional exceptions parameters e in the rule phase instead of only one for PROB to account for this finding.

The third competitor model, referred to here as KL, is based on *causal Bayes nets* (Pearl, 2000) and implements an idea by Hartmann and Rafiee Rad (2012; see also Paris, 1998). The updating mechanism is similar to that in PROB, introducing a new and smaller exceptions parameter e' for each content in the rule phase, but the other model parameters are also adjusted in going from knowledge phase to rule phase so that the so-called Kullback-Leibler distance between the probability distributions parameterized by the parameters for each phase is as small as possible.

Table 5 provides an overview of the data sets used in the meta-analysis and the number of free parameters that the DSM and the competitor models need to fit an individual participant. As noted by many authors (e.g., Roberts & Pashler, 2000), model comparison needs to take model fit and model flexibility into account. This is frequently done by means of model-selection indices such as AIC and BIC that penalize model flexibility in terms of the number of free parameters. Because we do not have the statistical apparatus of maximum-likelihood estimation on which these methods rely at our disposal here, we rely only on model fit. Note, however, that the DSM invariably requires at most as many parameters as, and often fewer parameters than, its competitors (see Table 5). The results presented below are thereby stacked against the DSM because explicit penalties for model flexibility based on the number of free parameters would hurt

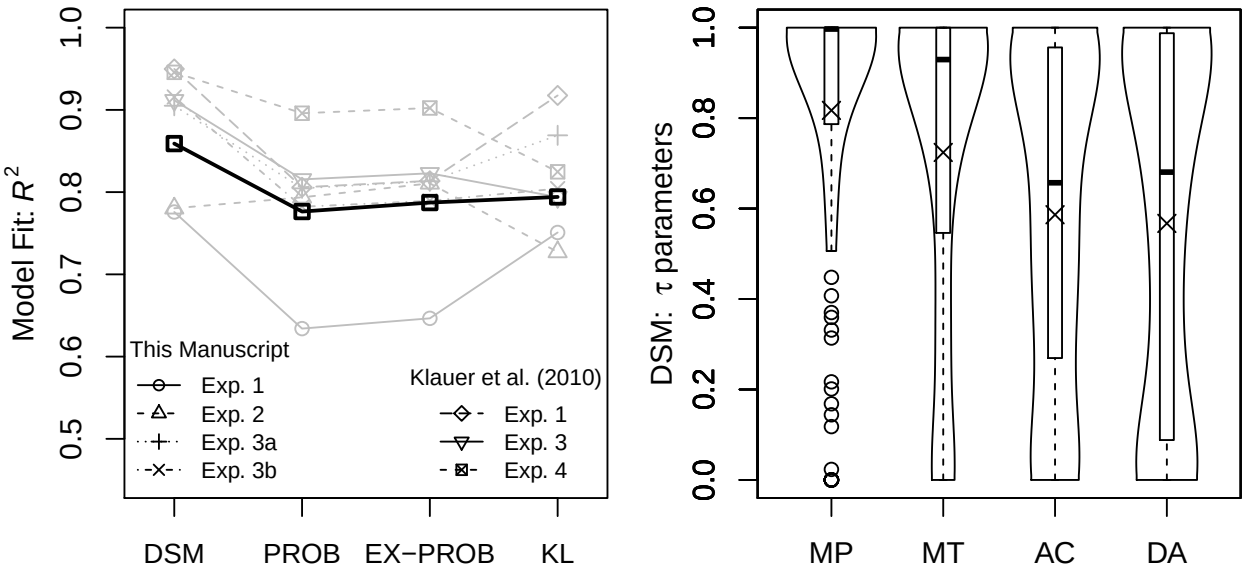

Figure 5. The left panel shows mean model fits (R^2) of each experiment (in grey) and the weighted grand mean (in black) of the four candidate models. The right panel shows violin-plots for the τ parameters of the DSM across all data sets and contents.

the DSM less severely than the competitor models.

We submitted the R^2 goodness-of-fit values of the four models fitted to each participant's data to a LMM with random effects for participants and experiment. This allowed us to assess which model provides the best account while taking into account the idiosyncrasies of each study (see e.g., Singmann, Klauer, & Kellen, 2014, for a similar approach). We set up the LMM in such a way that participants were weighted equally across studies (as traditionally done in meta-analyses; Hedges & Olkin, 1985) and then assessed the overall effect of model (DSM, PROB, EX-PROB, vs. KL). More details on the data sets used, the exact specification of the different Bayesian models, and the specification of the LMM can be found in the online appendix.

Results and Discussion

Mean model fits for the different experiments and models as well as the grand mean for each model are displayed in Figure 5 (left panel). As can be seen, the four models differ in their ability to fit the data, with the DSM consistently providing a better account than the competitor models. This was confirmed by the LMM which revealed a significant main effect of model, $\chi^2(3) = 17.66$, $p = .0005$. The DSM provided a better account than each of the three competitor models, smallest $z = 3.63$, largest $p_F = .0008$, which did not differ significantly from each other, largest $|z| = 1.35$, smallest $p_F = .34$. In other words, the DSM explained additional variance over and above the Bayesian competitor models, mean $R^2 = .86$ versus mean $R^2 = .79 - .80$.[7]

Thus, the effect of the conditional does not seem to be to alter the knowledge base from which individuals reason. Rather, it provides the reasoner with a content-independent form-based information on the (subjective) acceptability of the inference which is integrated with the information provided by one's background knowledge.

The Bayesian models incorporate different versions of updating by a conditional rule. While all of these instanstiations of updating performed equally in terms of goodness of fit, it is possible that yet another version of updating could be found that outperforms the DSM. Another possible limitation of the Bayesian models tested here is that they assume certainty of the minor premise (i.e., that the probability of the minor premise is 1). However, Singmann, Klauer, and Over (2014) have shown that Bayesian models not employing this restriction (e.g., Pfeifer & Kleiter, 2010) also do not provide a descriptively adequate account. More importantly, it was not clear to us if and how such models could be fitted to the data presented here.

As noted previously, one general prediction of the DSM is that we expected the pattern of form-parameters τ to mimic the pattern usually found with abstract materials: MP > MT > AC ≥ DA. We also performed a meta-analysis on the τ parameters across the seven data sets to evaluate this prediction. The distribution of τ parameters across all participants and items is shown in Figure 5 (right panel) and is descriptively in

[7]Note that when analyzing the data sets separately, the DSM provided a significantly better account than all three competitor models in only two of the seven data sets. At the same time, in none of the data sets did a competitor model provide a significantly better account than the DSM.

line with the prediction. The τ parameters were submitted to an LMM equivalent to the one reported for the R^2 values, but with fixed effects for inference (MP, MT, AC, vs. DA) instead of model. This LMM revealed a significant main effect of inference, $\chi^2(3) = 12.91$, $p = .005$. Follow-up contrasts (using one-sided tests for the first two comparisons) confirmed that MP was larger than MT by $M = .09$, $z = 2.20$, $p_H = .02$, MT was larger than AC and DA by $M = .15$, $z = 2.47$, $p_H = .02$, and there was no difference between AC and DA, $M = .02$, $z = .49$, $p_H = .62$.

General Discussion

A dual-source model of probabilistic conditional reasoning was investigated. The tenability of its assumptions was empirically assessed in terms of model fit, a model-comparison exercise, a non-parametric test of a unique prediction, and selective-influence studies. In terms of model fit, the DSM accounted on average for $R^2 = 86\%$ of the variance in reasoners' endorsements. This exceeds the amount accounted for by three competitor models. The knowledge-based component, which is an instantiation of the Bayesian model by Oaksford et al. (2000), explains most of the variance, 55%, when fitted individually, while the form-based component alone accounts for only 5%. Interestingly, the DSM thereby explains substantially more variance than its two subcomponents taken together. This shows that the weighted interplay of form-based and knowledge-based component is crucial for providing an adequate account of the data. Adequate model fit implies that the assumptions characterizing the model are satisfied by the data to a reasonable first degree of approximation.

Going beyond model fit our model selection exercise shows that the DSM provides a better account then three purely Bayesian competitor models while employing on average fewer parameters. This provides further evidence for the idea of a content-independent form-based process on top of the knowledge-based part shared by all surveyed models. The non-parametric test of a unique prediction shows that our assumption of how reduced and full inferences are related – namely such that the ratio in Equation 8 is not a function of content – holds for the observed endorsements irrespective of any model fits. Importantly, this unique prediction goes beyond a redescription of the empirical observation that endorsements increases from reduced to full inferences and is not predicted by the Bayesian models.

Finally, selective-influence studies are motivated by the fact that the parameters of a model are initially only mathematical entities devoid of psychological meaning. Such studies seek to validate the psychological interpretation attached to the parameters. For example, if our "knowledge parameters" ξ model reasoners' background knowledge then they should respond to variations between contents in the availability of counterexamples, of disablers and alternative causes in specific patterns, which they generally did (Exp. 1; see also Exp. 3). Similarly, a manipulation of argument form (replacing conditional premises with biconditional premises) should selectively affect the form-based parameters τ (Experiment 1), and a manipulation of speaker expertise should selectively affect the weight λ given to the rule and all inferences based on it (Exp. 2), predictions that were satisfied by the data.

The DSM and its Relationship to Dual-Process Models

In some respects, the DSM is similar to traditional dual-process and dual-system models (Evans & Stanovich, 2013) that postulate that two qualitatively different processes or systems shape human cognition. It is instructive to review criticisms of dual-process models as listed by Keren (2013) to see (a) whether the DSM can be defended against them and (b) to bring in sharper relief commonalities and differences of the DSM and current dual-process models, and (c) to accentuate what is new in the DSM (see also Beller & Spada, 2003, pp. 365-367).

One criticism of traditional dual-process models is that one-process models can often provide parsimonious alternative accounts (e.g., Keren, 2013). In the present context, a natural one-process competitor is a Bayesian account extended by an updating mechanism for updating by a conditional rule. As already discussed, there is no agreed-upon or normatively distinguished way of updating by a conditional in the Bayesian framework, but we considered three descriptive Bayesian models that instantiate different plausible possibilities of updating by a conditional. In a model-comparison exercise across seven studies, the DSM outperformed all three Bayesian models. Nevertheless it is possible that an alternative instantiation of updating could be found that provides a better account of the data than the DSM in which case the Bayesian model would be preferred as conceptually more simple.

From a higher vantage point on this issue, it may also be possible to re-conceptualize the DSM as a Bayesian model noting that the mixing of two response proposals, one knowledge-based, the other form-based, could be seen as an instance of model averaging in a Bayesian framework. Indeed, the knowledge-based component is already a Bayesian model, and the remaining challenge here would be to model the form-based component in a Bayesian framework, perhaps along the lines exemplified by von Sydow (2011) in his Bayesian logic.

As argued by Keren (2013), further problems of dual-process models stem from the fact that they characterize the different processes in terms of attributes such as whether or not working memory, intention, effort, and so on are involved. In contrast, the DSM defines its two components *operationally* in a well-defined empirical paradigm, providing measures of them via the parameter estimates. This then al-

lows one to study empirically how the different components react to manipulations of working memory resources, intentions, effort, and so forth.

In Keren's (2013) view, dual-process models furthermore do not state explicit constraints, and consequently do not generate testable predictions. In contrast, the DSM imposes specific restrictions on the data that we worked out in Equation 8. These restrictions were tested directly and the results were in line with the predictions. Moreover, beyond the basic paradigm, testable predictions were generated and tested in the selective-influence studies as was done for the reanalysis of data sets by Markovits et al. (2015).

On the basis of the validation of the model parameters via selective-influence studies, the model was applied to dissect classical suppression effects into effects on the knowledge-based and the form-based component as well as to reanalyze recent studies by Markovits et al. (2015). Suppression effects of disablers and alternatives were found to differ qualitatively. While both kinds of counterexamples suppressed the form-based component in meaningful and analogous ways, alternatives exerted a more pronounced effect on the knowledge-based component than disablers. This suggests that disablers are spontaneously taken into account in probabilistic assessments of causal contents, whereas alternatives need to be made explicit to obtain adequate consideration. These findings are reminiscent of a neglect of alternative causes in causal reasoning recently reported by Fernbach and Erb (2013), and provide a new insight in how suppression effects are generated. The same is true of the observation that the weight given to the formal component, the conditional rule and the inferences based on it, is depressed by disablers, but not by alternatives, providing an empirical justification for seeing effects of speaker (in)expertise and of disablers both as instances of the same suppression effect as is sometimes done in the literature (e.g., Evans & Over, 2004, p. 106).

Turning back to Keren's (2013) list, his criticism that dual-process models do not generate new insights or questions can thus not be maintained for the DSM. Yet another problem of dual-process models as seen by Keren (2013) is that the assumption of discrete processing types is undermined by the demonstration of processing continua. Our reanalysis of Markovits et al.'s (2015) indeed suggests that the deductive and probabilistic instructions contrasted in these studies differ by differential weighting of the form-based and knowledge-based contributions and thus, along a continuum defined by the weighting parameter rather than in terms of discrete processing types. In other words, the DSM accommodates evidence for processing continua in terms of the continuum of mixtures of two processing types with continuously differing weight λ.

Taken together, the DSM does not support the idea that probability theory can function as a wholesale replacement for logic as a computational-level theory of what inferences people should draw (Chater & Oaksford, 2001). Instead, it suggests that some ideas from the old paradigm capitalizing on the form of presented arguments (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Rips, 1994) need to be transferred to the new paradigm. The DSM provides a tentatively validated tool for disentangling knowledge-based and presumably probabilistic components from form-based components remaining as agnostic as possible on the attributes and characteristics of these two components, permitting the data to resolve such questions.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*, 158–175. doi:10.3758/s13428-011-0123-7

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379–384. doi:10.3758/BF03192707

Baratgin, J. & Politzer, G. (2006). Is the mind Bayesian? The case for agnosticism. *Mind & Society*, *5*, 1–38. doi:10.1007/s11299-006-0007-1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001

Batchelder, W. H. & Alexander, G. E. (2013). Discrete-state models: Comment on Pazzaglia, Dube, and Rotello (2013). *Psychological Bulletin*, *139*, 1204–1212. doi:10.1037/a0033894

Beller, S. (2008). Deontic norms, deontic reasoning, and deontic conditionals. *Thinking & Reasoning*, *14*, 305–341. doi:10.1080/13546780802222258

Beller, S. & Kuhnmünch, G. (2007). What causal conditional reasoning tells us about people's understanding of causality. *Thinking & Reasoning*, *13*, 426–460. doi:10.1080/13546780701319122

Beller, S. & Spada, H. (2003). The logic of content effects in propositional reasoning: The case of conditional reasoning with a point of view. *Thinking & Reasoning*, *9*, 335–378. doi:10.1080/13546780342000007

Bonnefon, J.-F. & Politzer, G. (2010). Pragmatic conditionals, conditional pragmatics, and the pragmatic component of conditional reasoning. In M. Oaksford & N. Chater (Eds.), *Cognition and Conditionals: Probability and Logic in Human Thought* (pp. 233–249). Oxford: Oxford University Press.

Bretz, F., Hothorn, T., & Westfall, P. H. (2011). *Multiple comparisons using R*. Boca Raton, FL: CRC Press.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 61–83. doi:10.1016/0010-0277(89)90018-8

Byrne, R. M. J. (1991). Can valid inferences be suppressed? *Cognition*, *39*, 71–78. doi:10.1016/0010-0277(91)90060-H

Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, *40*, 347–373. doi:10.1006/jmla.1998.2622

Chan, D. & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, *53*, 217–238. doi:10.1016/0010-0277(94)90049-3

Chater, N. & Oaksford, M. (2001). Human rationality and the psychology of reasoning: Where do we go from here? *British Journal of Psychology*, *92*, 193–216. doi:10.1348/000712601162031

Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 811–823. doi:10.1002/wcs.79

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Bayesian reasoning with ifs and ands and ors. *Cognition*, *6*, 192. doi:10.3389/fpsyg.2015.00192

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, *23*, 646–658. doi:10.3758/BF03197265

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274–282. doi:10.3758/BF03211151

Evans, J. S. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, *48*, 1–20. doi:10.1016/0010-0277(93)90056-2

Evans, J. S. B. T. & Over, D. E. (2004). *If*. OUP Oxford.

Evans, J. S. B. T. & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241. doi:10.1177/1745691612460685

Fernbach, P. M. & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1327–1343. doi:10.1037/a0031851

Geiger, S. M. & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition*, *35*, 2060–2074. doi:10.3758/BF03192938

Haigh, M., Stewart, A. J., & Connell, L. (2013). Reasoning as we read: Establishing the probability of causal conditionals. *Memory & Cognition*, *41*, 152–158. doi:10.3758/s13421-012-0250-0

Hartmann, S. & Rafiee Rad, S. (2012). *Updating on conditionals = Kullback-Leibler + causal structure*. Unpublished MCMP working paper, München.

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience* (pp. 25–48). DOI: 10.1007/978-1-4939-2236-9_2. Springer New York.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

Hintze, J. L. & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician*, *52*, 181–184. doi:10.2307/2685478

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

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

Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction.* Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N. & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678. doi:10.1037/0033-295X.109.4.646

Keren, G. (2013). A tale of two systems a scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*, 257–262. doi:10.1177/1745691613483474

Klauer, K. C., Beller, S., & Hütter, M. (2010). Conditional reasoning in context: A dual-source model of probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 298–323. doi:10.1037/a0018705

Lenth, R. V. (2015). *lsmeans: least-squares means*. R package version 2.20-1.

Liu, I.-m. (2003). Conditional reasoning and conditionalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 694–709. doi:10.1037/0278-7393.29.4.694

Markovits, H., Brisson, J., & de Chantal, P.-L. (2015). Deductive updating is not Bayesian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 949–956. doi:10.1037/xlm0000092

Matarazzo, O. & Baldassarre, I. (2008). Probability and instruction effects in syllogistic conditional reasoning. *Proceedings of World Academy of Science, Engineering and Technology*, *2*, 375–383.

Neth, H. & Beller, S. (1999). How knowledge interferes with reasoning - suppression effects by content and context. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of*

the 21st Annual Conference of the Cognitive Science Society (pp. 468–473). Mahwah, N.J.: Erlbaum.

Nickerson, R. S. (2015). *Conditional reasoning: the unruly syntactics, semantics, thematics, and pragmatics of "if"*. New York: Oxford University Press.

Oaksford, M. & Chater, N. (2003). Computational levels and conditional inference: Reply to Schroyens and Schaeken (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 150–156. doi:10.1037/0278-7393.29.1.150

Oaksford, M. & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford; New York: Oxford University Press.

Oaksford, M. & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, *19*(3-4), 346–379. doi:10.1080/13546783.2013.808163

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 883–899. doi:10.1037/0278-7393.26.4.883

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*, 431–438. doi:10.1080/13546780903266188

Paris, J. (1998). Common sense and maximum entropy. *Synthese*, *117*, 75–93. doi:10.1023/A:1005081609010

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, U.K.; New York: Cambridge University Press.

Pfeifer, N. & Kleiter, G. D. (2010). The conditional in mental probability logic. In M. Oaksford & N. Chater (Eds.), *Cognition and Conditionals: Probability and Logic in Human Thought* (pp. 153–173). Oxford: Oxford University Press.

Politzer, G. (2003). Premise interpretation in conditional reasoning. In D. Hardmann & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment, and decision making* (pp. 77–94).

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, Mass.: MIT Press.

Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367. doi:10.1037/0033-295X.107.2.358

Schroyens, W. J. & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 140–149. doi:10.1037/0278-7393.29.1.140

Singmann, H. & Klauer, K. C. (2011). Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & Reasoning*, *17*, 247–281. doi:10.1080/13546783.2011.572718

Singmann, H., Klauer, K. C., & Kellen, D. (2014). Intuitive logic revisited: New data and a bayesian mixed model meta-analysis. *PLoS ONE*, *9*, e94223. doi:10.1371/journal.pone.0094223

Singmann, H., Klauer, K. C., & Over, D. E. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, *5*, 316. doi:10.3389/fpsyg.2014.00316

Stenning, K. & van Lambalgen, M. (2010). The logical response to a noisy world. In M. Oaksford & N. Chater (Eds.), *Cognition and Conditionals: Probability and Logic in Human Thinking* (pp. 85–102). Oxford: Oxford University Press.

Stevenson, R. J. & Over, D. E. (2001). Reasoning from uncertain premises: Effects of expertise and conversational context. *Thinking & Reasoning*, *7*, 367–390. doi:10.1080/13546780143000080

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, *22*, 742–758. doi:10.3758/BF03209259

von Sydow, M. (2011). The Bayesian logic of frequency-based conjunction fallacies. *Journal of Mathematical Psychology*, *55*, 119–139. doi:10.1016/j.jmp.2010.12.001