Cognition 150 (2016) 26–36

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The relevance effect and conditionals ☆,☆☆

Niels Skovgaard-Olsen [a,b,*], Henrik Singmann [c], Karl Christoph Klauer [b]

[a] Department of Philosophy, University of Konstanz, Konstanz, Germany
[b] Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
[c] Department of Psychology, University of Zürich, Zürich, Switzerland

ARTICLE INFO

Article history:
Received 25 August 2015
Revised 26 December 2015
Accepted 28 December 2015

Keywords:
Indicative conditionals
The New Paradigm
Relevance
Reasons
Concessive conditionals
The Equation

ABSTRACT

More than a decade of research has found strong evidence for P(if A, then C) = P(C|A) ("the Equation"). We argue, however, that this hypothesis provides an overly simplified picture due to its inability to account for relevance. We manipulated relevance in the evaluation of the probability and acceptability of indicative conditionals and found that relevance moderates the effect of P(C|A). This corroborates the Default and Penalty Hypothesis put forward in this paper. Finally, the probability and acceptability of concessive conditionals ("Even if A, then still C") were investigated and it was found that the Equation provides a better account of concessive conditionals than of indicatives across relevance manipulations.

1. Introduction

In philosophy, there is a widely shared consensus that Stalnaker's Hypothesis is wrong and that Adams' Thesis is correct, due to formal problems affecting the former but not the latter – known as the triviality results.

> Stalnaker's hypothesis: P(if A, then C) = P(C|A) for all probability distributions where P(A) > 0 and 'If A, then C' expresses a proposition.
> Adams' thesis: Acc(if A, then C) = P(C|A) for all simple conditionals (i.e., conditionals whose antecedent and consequent clauses are not themselves conditionals), where 'Acc(if A, then C)' denotes the degree of *acceptability* of 'If A, then C'.[1]
> Triviality Results: Lewis' triviality results show that there is no proposition whose probability is equal to P(C|A) for all probability distributions without the latter being subject to trivializing features such as that P(C|A) collapses to P(C) or that positive probabilities can only be assigned to two pairwise incompatible propositions (Bennett, 2003: chap. 5; Woods, 1997: chap. 4, p. 114–8).

In psychology, there has been a tendency to endorse a thesis very similar to Stalnaker's hypothesis, known as the Equation, which avoids the problems affecting the former by either denying that conditionals express propositions altogether or by endorsing three-valued de Finetti truth tables (Table 1).

At present, the theories united under the heading 'the New Paradigm of Reasoning', which endorse the Equation, have branched out in different directions. To name just a few, in Baratgin, Politzer, and Over (2013) and Politzer, Over, and Baratgin (2010), the Equation is studied in relation to three-valued de Finetti truth tables in general and its relation to conditional bets is emphasized. In Pfeifer and Kleiter (2011) and Pfeifer (2013), the Equation is endorsed on the basis of a coherence-based probability logic that works with intervals of imprecise probabilities. However, what matters for our purposes is not so much the exact theory in which the Equation is embedded but rather the general commitment to the Equation. As it stands, over a decade of empirical research has found strong evidence in favor of the Equation and a recent study has begun to challenge Adams' Thesis, as nicely outlined in Douven (2015b: chap. 3, 4).

In contrast, a basic intuition that has emerged repeatedly throughout the history of philosophy is that in conditionals like 'If it rains, then the match will be cancelled' the antecedent and

☆ This work was supported by grants to Wolfgang Spohn and Karl Christoph Klauer from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program "New Frameworks of Rationality" (SPP 1516).
☆☆ Supplementary materials including all data and analysis scripts are available at: https://osf.io/j4swp/.
* Corresponding author at: Department of Philosophy, University of Konstanz, Konstanz, Germany.
E-mail addresses: niels.skovgaard.olsen@psychologie.uni-freiburg.de, n.s.olsen@gmail.com (N. Skovgaard-Olsen).
[1] One of the reviewers pointed out that Adams later abandoned this position in Adams (1998). To sidestep such exegetical issues we use the phrase 'Adams' Thesis' to denote the position attributed to him in the literature based on his earlier work.

http://dx.doi.org/10.1016/j.cognition.2015.12.017
0010-0277/

Table 1
De Finetti truth table.

A	C	If A, then C
⊤	⊤	⊤
⊤	⊥	⊥
⊥	⊤	*void*
⊥	⊥	*void*

The Equation: P(if A, then C) = P(C|A), but 'If A, then C' does not express a classical proposition (Bennett, 2003; Evans & Over, 2004; Oaksford & Chater, 2007).

the consequent should somehow be connected or relevant for one another as one aspect of the conditionals' meaning (for references see Douven, 2015b; Krzyżanowska, 2015; Skovgaard-Olsen, in press). This intuition is especially salient when we observe examples in which the relevance expectation is violated, as in conditionals such as 'If blood is red, then Oxford is in England', for which the truth-value of the antecedent leaves the truth-value of the consequent unaffected. However, surprisingly this intuitive idea is not preserved in any of the theories of conditionals currently endorsed in the psychology of reasoning, as we shall see.

1.1. The paradoxes of the material implication

Before the Equation became popular in the psychology of reasoning (Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer, 2013), the dominant theory was mental model theory, which is based on the material implication analysis of natural language conditionals (Johnson-Laird & Byrne, 2002).[2] Since the material implication is always true except in cases when its antecedent is true and its consequent is false, the theory validates the following argument-schemes that are known to give rise to nonsensical results once natural language content is substituted:

$$\frac{\neg A}{\therefore \text{if } A, C} \qquad \frac{C}{\therefore \text{if } A, C} \tag{1}$$

With no restrictions on the relationship between the antecedent and the consequent, *any* conditional could be inferred from a false antecedent or a true consequent, no matter how odd. Hence, from the true premise 'It is not the case that Europe has been ruled by France since Napoleon' the conditional 'If Europe has been ruled by France since Napoleon, then the sun emits light' could be inferred. And from the true premise 'The sun emits light', the conditional 'If Europe has been ruled by France since Napoleon, then the sun emits light', or indeed 'If Europe was liberated from occupation by Napoleon's France, then the sun emits light' could be inferred. Unsurprisingly, participants in psychological experiments tend to find such inferences odd as well (Pfeifer & Kleiter, 2011). Of course, this fact has not escaped the proponents of mental model theory. In accounting for the oddness of such inferences, they exploit the logical equivalence of the material implication with '¬A v C' and argue that the reason why we are reluctant to endorse the valid argument schemes in [1] is due to the problem with endorsing the following equally valid argument schemes:

$$\frac{\neg A}{\therefore \neg A \text{ v } C} \qquad \frac{C}{\therefore \neg A \text{ v } C} \tag{2}$$

Since more possibilities are excluded by the premises than by the conclusions in [2], information is lost in the conclusion, and according to Johnson-Laird and Byrne (2002; Byrne & Johnson-Laird, 2009) this is really the source of our intuitive problems with [1]. However, in the absence of a prior theoretical commitment to the logical equivalence of natural language conditionals to disjunctions, a much more straightforward diagnosis of the oddness of [1] runs as follows. The problem is not so much that *fewer* possibilities are excluded by the conclusion than by the premises, but rather that different conditions are imposed by the premises and the respective conclusions. The premises are silent on the relationship between A and C and impose conditions on a set of possible worlds by being factual propositions; the conclusions impose constraints on epistemic states (i.e., that A is epistemically relevant for C).

In contrast, the probabilistic approaches that are currently replacing the mental model theory under the heading 'the New Paradigm of Reasoning' endorse the Equation and reject [1]. Purportedly this is because the premises do not probabilistically constrain the conclusion when the latter is interpreted as a conditional probability as long as $0 < P(\text{premise}) < 1$ (Bennett, 2003, p. 139; Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer & Kleiter, 2011). However, as argued in Skovgaard-Olsen (in press), it can be claimed that these theories reject [1] for the wrong reasons. The most obvious diagnosis of the oddness of [1] remains that no restrictions on the relevance of A for C are introduced by the premises, whereas indicative conditionals fit for the speech act of assertions seem to require A to be relevant for C. Yet these probabilistic approaches within the New Paradigm of Reasoning are unable to account for this. According to the latter, indicative conditionals should be seen as a linguistic device by which the participants activate a mental algorithm known as the Ramsey test, which consists in temporarily adding the antecedent to their knowledge-base and evaluating the consequent under its supposition (Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer, 2013). As such, indicative conditionals can have a high probability of being true as long as P(C) is high, even if the antecedent is irrelevant for the consequent. Accordingly, none of the main contenders in contemporary psychological accounts of conditional reasoning are willing to make relevance part of the core meaning of natural language conditionals.[3]

1.2. P(If A, then C) and relevance

The next surprise is that until quite recently,[4] when the role of relevance in the interpretation of conditionals was empirically investigated it was either found that no support could be provided (Oberauer, Weidenfeld, & Fischer, 2007; Singmann, Klauer, & Over, 2014), or that it was only weakly supported by the data (Over et al., 2007). So perhaps relevance should be set aside for our theories of conditionals after all. In these studies, relevance was operationalized in terms of the Δp rule, which is well-known from the psychological literature on causation, where $\Delta p > 0$ has been taken to be a necessary, but not sufficient, condition for inferring causality (Cheng, 1997).

The Δp Rule: $\Delta p = P(C|A) - P(C|\neg A)$

As P(C|A) is already occupied as a predictor of P(if A, then C) by the Equation, Over et al. (2007) and Singmann et al. (2014) try to obtain an orthogonal predictor for the relevance approach by using P(C|¬A). The evidence clearly favored P(C|A) as a predictor.

[2] As one of the reviewers pointed out, mental model theory has recently been revised so as to avoid being committed to the material implication analysis of the natural language conditional in Johnson-Laird, Khemlani, and Goodwin (2015). However, we here restrict our focus to the previous version of the theory.

[3] However, it should be noted that Over and Evans (2003) did entertain the possibility that relevance could characterize a subgroup of conditionals (i.e. causal conditionals). Yet this idea was later rejected in Over, Hadjichristidis, Evans, Handley, and Sloman (2007).

[4] An exception is Douven, Elqayam, Singmann, Over, and Wijnbergen-Huitink (in press). In this study it was found in a novel experimental task that the participants used clues about the inferential relations between A and C in evaluating the conditionals used in that task.