Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-025-02725-2

BRIEF REPORT

Probabilities of conditionals: The relevance effect might be confounded by the existence of boundary cases

Likan Zhan[1,2,3] · Meng Wang[4]

Accepted: 3 June 2025

Abstract
The relevance between antecedent and consequent has recently been regarded as essential in modulating the probability assigned to a conditional in natural language. The empirical results are mixed. Positive evidence mainly comes from intensional probability studies about ordinary, unique events. Extensional probability studies about novel abstract events commonly fail to observe such an effect. In extensional probability studies, a set of events is typically provided to sustain participants' judgments. Depending on whether the antecedent and the consequent are true or false, the set can be divided into four subsets. When one or more subsets are empty, the set is called a boundary case. When the number of events becomes smaller, it becomes easier for boundary cases to occur. In previous extensional probability studies, however, boundary cases were normally not included in the test stimuli. In intensional probability studies, no explicit events are provided; participants have to mentally simulate a set of events from their own background knowledge to help them make judgments. The size of the mentally simulated sample is relatively small, especially when the judged statements are complex, like conditionals. It is then highly probable for the intensional probability studies to contain boundary cases, even though they cannot be directly observed. Based on the previous analyses, we suspect that the difference observed in previous studies might be confounded by the fact that boundary cases were included in the former case but not in the latter. To test this possibility, we introduced boundary cases into our experiment involving abstract multiple events and observed that (1) when boundary cases were included in the analyses, modulation effect was observed for three of the four parameters; (2) when boundary cases were excluded from analyses, no modulation effect was observed. Reanalyses of previous intensional studies corroborated our hypothesis. We also discussed the potential reason why relevance effect and boundary cases cooccur.

Keywords Probabilities of conditionals · Conditional Probability · Extensional versus intensional probability · The Relevance effect · Boundary cases

Introduction

A compound statement like "If A then C" is called a conditional, with proposition *A* called the antecedent, and proposition *C* called the consequent. To determine the probability of indicative conditionals in natural language is essential to differentiate between different theories of conditionals.

Probability of conditional and conditional probability

Different proposals have been put forward in the literature regarding the probability assigned to the conditional (Over & Evans, 2024). First, in the earlier stage, experiments were typically conducted to determine whether the material implication (Frege, 1960; Russell, 1906) defined in Boolean logic could explain the probability assigned to

✉ Likan Zhan
zhanlikan@hotmail.com

Meng Wang
wm0420@126.com

1 Cognitive Science and Allied Health School, Beijing Language and Culture University, Beijing 100083, China

2 Institute of Life and Health Sciences, Beijing Language and Culture University, Beijing 100083, China

3 Key Laboratory of Language and Cognitive Science (Ministry of Education), Beijing Language and Culture University, Beijing 100083, China

4 Graduate School, Beijing Language and Culture University, Beijing 100083, China

the conditional (material conditional theory). According to the material conditional theory (Jeffrey, 2006), the conditional "If A then C" is true in any case except when A is true and C is false. The probability of the conditional then equals $1 - p(A\neg C)$, or equivalently to the sum of $p(AC)$, $p(\neg AC)$, and $p(\neg A\neg C)$ (see Appendix A for details on how these terms are defined). This viewpoint has primarily been disconfirmed from studies using abstract frequency information (Evans et al., 2003; Oberauer & Wilhelm, 2003) and studies using ordinary unique situations (Over et al., 2007; Singmann et al., 2014). Second, the suppositional theory (Adams, 1965; Edgington, 1995; Oaksford & Chater, 2020; Stalnaker, 1970) hypothesized that the probability of a conditional is the same as the conditional probability $p(C|A)$. Third, the mental model theory (Johnson-Laird et al., 1999; Lopez-Astorga et al., 2021) proposed that cases where the antecedent A is false have no bearing on whether the conditional "If A then C" is true or false. They have no bearing on the conditional's probability either. The probability of the conditional either equals the joint probability $p(AC)$, or equals to the conditional probability $p(C|A)$.

The empirical literature has identified the conditional probability $p(C|A)$ as the main predictor of the probability of conditionals, both in studies using abstract frequency information (Evans et al., 2003; Fugard et al., 2011; Girotto & Johnson-Laird, 2004; Oberauer & Wilhelm, 2003; Oberauer et al., 2007) and in studies using ordinary unique situations (Over et al., 2007; Singmann et al., 2014; Skovgaard-Olsen et al., 2016, 2019). Not being orthogonally independent of the conditional probability, the joint probability $p(AC)$ has also been observed as a significant predictor in some studies (Evans et al., 2003; Fugard et al., 2011; Girotto & Johnson-Laird, 2004; Oberauer & Wilhelm, 2003; Oberauer et al., 2007).

The relevance effect

In recent years, the research focus has shifted to determine whether the conditional probability is sufficient to explain the probability of conditionals. It is believed that for a conditional to be accepted as true, there has to be an inferential link from the antecedent to the consequent, normally called the *relevance effect* (Skovgaard-Olsen et al., 2016). The inferential relation between the antecedent and the consequent can be inductive, abductive, or deductive (Douven et al., 2018). The current study focuses on four variations of such deductive parameters.

First, some researchers (White, 2003) hypothesized that for a conditional to be assertable, the probability of confirmatory instances should be larger than that of disconfirmatory instances. Given the conditional "If the card is red, then it is round," the cards that are red and round, together with the cards that are neither red nor round, are called confirmatory instances. On the contrary, the cards that are red but not round, together with those that are not red but round, are called disconfirmation instances. The modulating parameter is the probability difference between the confirmatory and disconfirmatory instances (Table 1, $\Delta p1$). Second, others (Crupi & Iacona, 2021; Douven, 2015) held that the conditional probability $p(C|A)$ should be larger than the marginal probability $p(C)$. The modulating parameter is then defined as the probability difference between conditional probability $p(C|A)$ and the marginal probability $p(C)$, weighted by the marginal probability $p(\neg C)$ (Table 1, $\Delta p2$). Third, still others (Rott, 1986, 2019; Spohn, 2012) believed that the conditional probability $p(C|A)$ should be larger than the conditional probability $p(C|\neg A)$. The modulating parameter is then defined as the probability difference between $p(C|A)$ and $p(C|\neg A)$ (Allan, 1980; Shanks, 1995; Sheps, 1958; Table 1, $\Delta p3$) or the probability difference between $p(C|A)$ and $p(C|\neg A)$, weighted by the probability $1 - p(C|\neg A)$ (Cheng, 1997; van Rooij & Schulz, 2018; Table 1, $\Delta p4$).

One specific account that directly connects the modulation parameter $\Delta p3$ to the conditional probability is the default-penalty hypothesis (Skovgaard-Olsen et al., 2016, 2019). According to this hypothesis, when the relevance as described by $\Delta p3$ is positive, the probability of the conditional will equal the conditional probability. When the relevance is not positive, however, the equality will be disrupted. Empirically, the predictability of the conditional probability will be positively related to the modulation parameter $\Delta p3$: The effect of the conditional probability on the judged probability of the conditional will be weakened as the modulation parameter $\Delta p3$ becomes smaller.

Extensional probability and intensional probability

The empirical results about relevant effects are mixed. Positive evidence mainly comes from studies using realistic stimuli (Douven et al., 2022; Krzyżanowska et al., 2017; Over et al., 2007; Skovgaard-Olsen et al., 2016, 2017, 2019). In these studies, propositions are usually chosen to describe some ordinary situation that participants have almost surely

Table 1 Parameters to modulate the probability of conditional (see Appendix A for details on how these terms are defined)

$$\Delta p_1 = [p(AC) + p(\neg A\neg C)] - [p(\neg AC) + p(A\neg C)]$$
$$\Delta p_2 = \frac{p(C|A) - p(C)}{p(\neg C)} = \frac{p(C|A) - p(C)}{1 - p(C)}$$
$$\Delta p_3 = p(C|A) - p(C|\neg A)$$
$$\Delta p_4 = \frac{p(C|A) - p(C|\neg A)}{1 - p(C|\neg A)} = \frac{p(C|A) - p(C)}{p(\neg A\neg C)}$$

experienced in their daily lives before the experimental setting. The situation is typically unique. In this case, the probability is simulated from the participant's personal knowledge before the experimental setting (Costello & Watts, 2014). These probabilities are called intensional probabilities (Johnson-Laird et al., 1999; Lopez-Astorga et al., 2021; Tversky & Kahneman, 1983), because personal knowledge is subjective, independent of the experimental setting, and can differ between participants. Intensional probabilities cannot be directly calculated from the experimental setting and must be indirectly retrieved from participants' subjective judgment.

Negative evidence mainly results from studies using artificial stimuli (Oberauer et al., 2007; Singmann et al., 2014). In these studies, propositions usually describe abstract materials of which participants have no background knowledge. A set of new situations is provided explicitly or in the form of frequency information, with which the probabilities could be deduced. In this case, the probabilities can be directly calculated from the ratio or the relevant frequencies of the cards provided in the experimental setting. These probabilities are called extensional probabilities (Johnson-Laird et al., 1999; Lopez-Astorga et al., 2021; Tversky & Kahneman, 1983) because they are objective and are supposed to be the same for all participants.

Experiment preview

The crucial aspect distinguishing the two groups of studies is that the information used to estimate the probability of the conditional is provided in extensional probability studies but not in intensional probability studies.

In extensional probability studies, a set of cards is typically provided to help participants' probability judgment of the conditional "if the card is red then it is round." Depending on whether the antecedent and the consequent are true or false, the set of cards can be divided into four subsets. When the number of cards in the given set becomes smaller, it usually gets easier for one or more subsets to become empty. In previous studies, however, the four subsets were artificially manipulated to contain at least one card; that is, none of the subsets was empty.

In intensional probability studies, no information is provided to help participants' probability judgments. To judge the probability of the conditional statement, participants have to retrieve a series of episodes from the long-term memory of their own background knowledge and mentally simulate the sampling process similar to that described in the extensional studies (Costello & Watts, 2014; Zhu et al., 2020). Judging a complex statement (a conditional statement is complex) is computationally expensive, and only a small number of cases can be simulated. It is then highly probable that one or more subsets of the mentally simulated episodes are empty, and one or more joint probabilities are judged to be zero. In the probabilistic truth-table task reported by Douven et al. (2022), for example, 63.9% of trials (3,610 of 5,650 trials) have one or more joint probabilities equaling zero (cf.: https://osf.io/7x63j/). Joint probabilities are not available in the experiments of Skovgaard-Olsen et al. (2016). However, some clues can be used to estimate the proportion. For example, $p(C|A)=0$ means that the joint probability $p(AC)=0$ and the subset AC is empty, and $p(C|A)=1$ means that the joint probability $p(A\neg C)=0$ (i.e., the subset $A\neg C$ is empty). When the judged probability of the conditional is 0 or 1, it is also highly probable (not necessarily) that one or more joint probabilities are zero. Based on these considerations, we suspect that about 34.4% of trials (756 of 2,196 trials) in Skovgaard-Olsen et al. (2016) contain at least one empty subset (cf.: https://osf.io/j4swp/).

Based on the previous analyses, we suspect that the difference between extensional and intensional probability studies may be confounded by whether empty subsets were included in the experimental settings. To tease this confounding apart, we conducted an extensional probability study, including boundary conditions with empty subsets.

Our experiment

Participants

Eighty college students (33 men, 47 women) from Beijing Language and Culture University, with an average age of 22.56, participated in the experiment. All participants were native Chinese, had not studied any logic course before, and had not participated in related experiments. Informed consent was obtained from each participant, and they were debriefed about the study's aims after completing the experiment. This study was conducted in accordance with the Beijing Language and Culture University Committee on Human Research Protection's recommendations, with written informed consent obtained from all subjects. The Beijing Language and Culture University Committee on Human Research Protection approved the protocol.

Stimuli

Each trial consists of a test sentence printed at the top and a test image at the bottom (Fig. 1). The test sentence remains the same across all test trials (i.e., "If the card is red, then it is round"; this sentence is the English translation of the original Chinese sentence). A test image presents a set of cards. Cards used in all test images have only two possible colors (Red and Blue) and two possible shapes (Round and Square). The set of cards in a trial was composed of four subsets: AC (Red–Round), $A\neg C$ (Red–Square), $\neg AC$

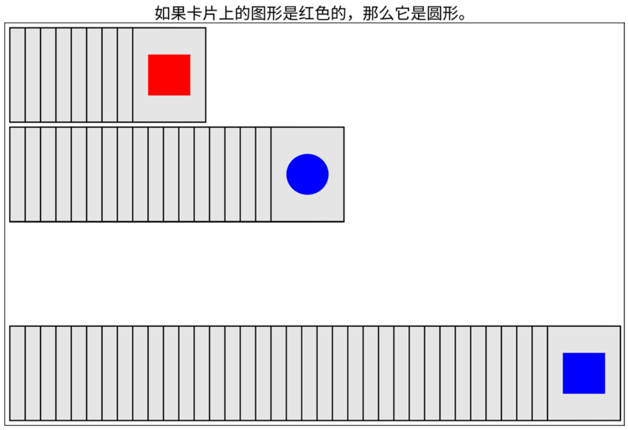

Fig. 1 An example of the test stimuli. (Color figure online)

(Blue–Round), and ¬*A*¬*C* (Blue–Square). Each subset of cards was displayed as a separate row in the test image: Cards in the same subset were placed in the same row, and cards from different subsets were placed in different rows. Cards in each row were partially overlapped, allowing participants to see only the far-right card content. However, the number of cards in a subset was indirectly reflected in the length of the corresponding row. The number of cards in each subset (each row) could be 0, 9, 18, or 36, resulting in 4 (*AC*: 0, 9, 18, 36) × 4 (*A*¬*C:* 0, 8, 18, 36) × 4 (¬*AC:* 0, 8, 18, 36) × 4 (¬*A*¬*C:* 0, 8, 18, 36) = 256 test images. A row with its length equaling the image's width indicates that the subset has 36 cards (Fig. 1, Row 4, ¬*A*¬*C*). An empty row suggests the subset has no card (Fig. 1, Row 3, *AC*). A row with a length between the previous two indicates that the subset has 9 (Fig. 1, Row 1, *A*¬*C*) or 18 (Fig. 1, Row 2, ¬*AC*) cards. The spatial orders of the four rows were pseudorandomly changed across the 256 images.

Procedure

The test stimuli were presented with PsychoPy 3.0 (Peirce et al., 2019). The experiment began with written instructions (in Chinese) that explained the test stimuli and participants' tasks. In each test trial, participants were asked to assume that the test sentence was uttered by someone who randomly selected a card from the given set without looking at the card's content. Using a probabilistic truth-table task, participants were asked to judge the probability for the test sentence "If the card is red, then it is round" to be true by entering a value between 0 (= absolutely impossible) and 100 (= definitely/absolutely sure) within 8 s. Participants were required to give their intuitive estimation without using a precise mathematical calculation. All 256 trials were presented to each participant in a pseudorandom order, with each trial offered on a separate page. Four more trials were presented as practice. It took approximately 20 min to complete the entire experiment.

Data processing

First, we divided the test stimuli into 2 (*AC*: empty vs. nonempty) × 2 (*A*¬*C*: empty vs. nonempty) × 2 (¬*AC*: empty vs. nonempty) × 2 (¬*A*¬*C*: empty vs. nonempty) = 16 groups according to whether a specific subset was empty or not. All groups except Group 16 contained one or more empty subsets. A histogram then displays the judged probabilities for each group (Fig. 2). The histograms enclosed in a red box labeled $n(AC) = 0$ represent the test stimuli for which the subset *AC* is empty. In contrast, the histograms enclosed in a blue box labeled $n(AC) \neq 0$ correspond to test stimuli where the subset *AC* was not empty. The histograms enclosed in the green box labeled $n(\neg A\neg C) = 0$ involve test stimuli for which the subset ¬*A*¬*C* is empty, while the histograms in the purple box labeled $n(\neg A\neg C) \neq 0$ involve test stimuli for which the subset ¬*A*¬*C* was not empty. The histograms in the columns labeled $n(A\neg C) = 0$ involved test stimuli where the provided cards did not include the red and square items; the histograms in the columns labeled $n(A\neg C) \neq 0$ involved test stimuli where the provided cards included the red and square items. The histograms in rows labeled $n(\neg AC) = 0$ involved test stimuli in which the provided cards did not include the blue and round ones. In comparison, the histograms in rows labeled $n(\neg AC) \neq 0$ involved test stimuli in which the provided cards included the blue and round ones.

Second, we filtered the obtained data to ensure that the explored modulation parameters were adequately defined before the statistical modeling (Table 2). (a) For $\Delta p1$ and the conditional probability $p(C|A)$ to be defined, $p(A)$ and $n(A)$ should not be zero (i.e., subsets *AC* and *A*¬*C* should not be empty at the same time). When conducting the analyses, Groups 1, 3, 9, and 11 were filtered out. (b) For $\Delta p2$ to be defined, $p(\neg C)$ and $n(\neg C)$ should not be zero (i.e., subsets *A*¬*C* and ¬*A*¬*C* should not be empty at the same time). When conducting the analyses, Groups 1, 3, 9, 11, 5, and 7 were excluded. (c). For $\Delta p3$ to be defined, $p(\neg A)$ and $n(\neg A)$ should not be zero (i.e., subsets ¬*AC* and ¬*A*¬*C* should not be empty at the same time). When conducting the analyses, Groups 1, 3, 9, 11, 2, 5, and 6 were excluded. (d). For $\Delta p4$ to be defined, joint probability $p(\neg A\neg C)$ and subset ¬*A*¬*C* should not be zero. When conducting the analyses, Groups 1, 3, 9, 11, 2, 5, 6, 4, 7, and 8 were excluded.

Third, we fitted the filtered data to a binomial generalized linear mixed model (GLMM) using the package *MixedModels.jl* (Alday & Bates, 2025) in the Julia (Bezanson et al., 2017) programming language to determine the modulation effect of the proposed parameters. The model used participants' probability ratings as the response variable. The model had two fixed terms: The conditional

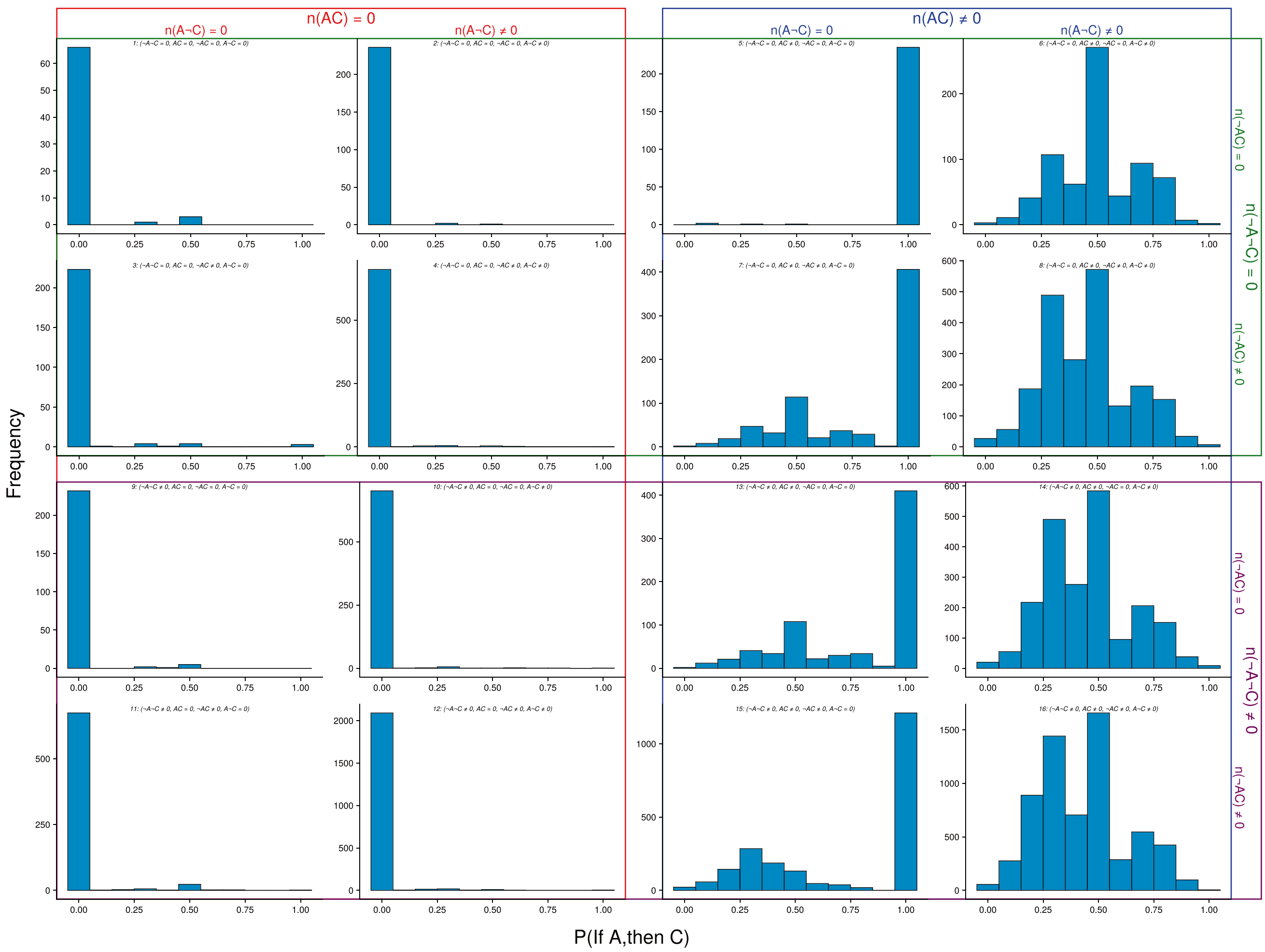

Fig. 2 Histograms of the 16 groups of the obtained data. (Color figure online)

Table 2 Criteria to filter the obtained data for specific modulation parameters

Parameter	$p(C \vert A)$	$p(C \vert \neg A)$	$p(\neg A \neg C)$	$p(\neg C)$	Groups Excluded
	$n(A) = n(AC) + n(A\neg C) = 0$	$n(\neg A) = n(\neg AC) + n(\neg A\neg C) = 0$	$n(\neg A\neg C) = 0$	$n(\neg C) = n(A\neg C) + n(\neg A\neg C) = 0$	
	$p(A) = p(AC) + p(A\neg C) = 0$	$p(\neg A) = p(\neg AC) + p(\neg A\neg C) = 0$	$p(\neg A\neg C) = 0$	$p(\neg C) = p(A\neg C) + p(\neg A\neg C) = 0$	
	Groups: 1,3,9,11	*Groups: 1,2,5,6*	Groups: 1,2,5,6;3,4,7,8	Groups: 1,3,5,7	
$\Delta P1$	×				1,3,9,11
$\Delta P2$	×			×	1,3,9,11; 5,7
$\Delta P3$	×	×			1,3,9,11; 2,5,6
$\Delta P4$	×	×	×		1,3,9,11; 2,5,6; 4,7,8

probability $p(C|A)$ and a specific modulation parameter being explored (Table 1). The fixed effects included the two terms' main effects and their interactions. The model had two random terms: Participants and items (images). The random effects included both the random intercept and the random slope for each random term. The model was written as $Probability \sim 1 + p(C|A) \times \Delta p + (1 + p(C|A) \times \Delta p \mid participants) + (1 + p(C|A) \times \Delta p \mid images)$ in *Julia*, where Δp stands for the specific modulation parameter being analyzed. The raw data, the Julia code, and the modal fitting's outputs were stored at the OSF repository (https://osf.io/t5xr3/?view_only=7bbb0782f44c4865a5479cc6fac254dd).

Results

Including versus excluding empty subsets

For each modulation parameter, we fitted the same model to two data sets: (a) all data were included in the analyses as long as the parameter was adequately defined (Fig. 2 and Table 2); (b) only data from Group 16 (Fig. 2) were included, where none of the four subsets was empty. The results were summarized in Table 3. The effects of $\Delta p1$ were different from the other three parameters. (1) For $\Delta p1$, the only significant effect was the conditional probability, regardless of the specific data set. (2) For parameters $\Delta p2$–$\Delta p4$, (a) when empty subsets were included in the analyses, both conditional probability and the modulation parameter, as well as their interaction, had significant effects; (b) when empty subsets were excluded from the analyses, the only significant predictor that survived was the conditional probability.

Comparing the four modulation parameters

To compare the effects of the four modulation parameters, we fitted the model *Probability ~ 1 + p(C|A) × Δp + (1|participants) + (1|images)* to the following two data sets: (a) The first data set includes empty subsets as long as all four parameters were appropriately defined. As a result, all data except for Groups 1, 3, 9, 11, 2, 5, 6, 4, 7, and 8 were included in the analyses. (b) The second data set excludes data with empty subsets, including Group 16 only. The results were summarized in Table 4. When empty subsets were included, it seemed that $\Delta p2$ was the best, followed by $\Delta p4$ and then $\Delta p3$. $\Delta p1$ was not as good as the other three parameters. When empty subsets were excluded from the analyses, the models for the four parameters were nearly identical.

Table 4 Results of model comparisons between different modulation parameters

Parameter	Loglikelihood	Deviance	AIC	AICc	BIC
Empty subsets included (Exclude groups: 1, 3, 9, 11, 2, 5, 6, 4, 7, 8)					
$\Delta P1$	−106833.0	166229.0	213678.0	213678.0	213723.0
$\Delta P2$	−106798.0	166159.0	213609.0	213609.0	213654.0
$\Delta P3$	−106810.0	166184.0	213633.0	213633.0	213678.0
$\Delta P4$	−106801.0	166165.0	213614.0	213614.0	213659.0
Empty subsets excluded (Only include group: 16)					
$\Delta P1$	−30272.0	29738.0	60556.0	60556.0	60597.0
$\Delta P2$	−30271.0	29737.0	60556.0	60556.0	60595.0
$\Delta P3$	−30272.0	29738.0	60556.0	60556.0	60596.0
$\Delta P4$	−30272.0	29738.0	60556.0	60556.0	60597.0

Extensional versus intensional probability

The dependent variable, the probability of the conditional, is bounded between 0 and 1. When the conditional probability *p(C|A)* equals 0 or 1, it indicates that empty subsets are present in the test stimuli. If the dependent variable is 0 or 1, it is also highly probable (though not sure) that empty subsets are included in the test stimuli. To compare our results with those obtained in intensional probability studies, we reanalyzed the results of Skovgaard-Olsen et al. (2016) with two changes: (a) In addition to the linear (mixed) model, we also fitted a generalized linear (mixed) model. (b) In addition to the full data, we also fitted the two models to the filtered data where the boundary cases were excluded: The cases where the judged conditional probability *p(C|A)* or the probability of the conditional was 0 or 1 (we must stress that the filtering criteria are rough, and the results are preliminary and tentative). The results (Table 5) were similar to our observations

Table 3 Effects of the conditional probability and the modulation parameter when empty subsets were included in or excluded from the analyses

Parameter	$p(C\|A)$	Δp	$p(C\|A) \times \Delta p$
Empty subsets included			
$\Delta P1$	$\beta = 4.31, z = 14.59, p < 1e\text{-}97$	$\beta = 0.26, z = 0.99, p = .3201$	$\beta = -0.48, z = -1.17, p = .2439$
$\Delta P2$	$\beta = 4.67, z = 19.08, p < 1e\text{-}80$	$\beta = 0.57, z = 3.14, p = .0017$	$\beta = -1.01, z = -3.30, p = .0010$
$\Delta P3$	$\beta = 4.16, z = 16.37, p < 1e\text{-}59$	$\beta = 0.79, z = 3.59, p = .0003$	$\beta = -1.29, z = -3.52, p = .0004$
$\Delta P4$	$\beta = 4.64, z = 15.97, p < 1e\text{-}56$	$\beta = 0.45, z = 3.34, p = .0009$	$\beta = -0.89, z = -3.44, p = .0006$
Empty subsets excluded			
$\Delta P1$	$\beta = 2.84, z = 12.70, p < 1e\text{-}36$	$\beta = 0.03, z = 0.09, p = .9289$	$\beta = 0.02, z = 0.03, p = .9738$
$\Delta P2$	$\beta = 2.88, z = 13.50, p < 1e\text{-}40$	$\beta = 0.13, z = 0.40, p = .6881$	$\beta = -0.22, z = -0.38, p = .7013$
$\Delta P3$	$\beta = 2.88, z = 12.31, p < 1e\text{-}34$	$\beta = 0.01, z = 0.04, p = .9719$	$\beta = -0.03, z = -0.05, p = .9635$
$\Delta P4$	$\beta = 2.87, z = 13.48, p < 1e\text{-}40$	$\beta = -0.01, z = -0.08, p = .9374$	$\beta = 0.03, z = 0.12, p = .9026$

Table 5 Reanalyses of Skovgaard-Olsen et al.'s (2016) experiment on indicative conditionals

Boundary Cases	$p(C\|A)$	Positive Relevance	Irrelevance	$p(C\|A)$: Positive Relevance	$p(C\|A)$: Irrelevance
Linear Mixed-Effects Model					
Include	$\beta = 0.62, z = 14.70, p < 1e\text{-}48$	$\beta = 0.18, z = 6.01, p < 1e\text{-}08$	$\beta = 0.01, z = 0.89, p = .3726$	$\beta = 0.15, z = 3.22, p = .0013$	$\beta = -0.19, z = -3.20, p = .0014$
Exclude	$\beta = 0.70, z = 15.76, p < 1e\text{-}48$	$\beta = 0.10, z = 3.66, p < 1e\text{-}55$	$\beta = 0.05, z = 2.13, p = .0330$	$\beta = 0.14, z = 2.89, p = .0039$	$\beta = -0.16, z = -2.32, p = .0202$
Generalized Linear Mixed-Effects Model					
Include	$\beta = 3.55, z = 11.14, p < 1e\text{-}28$	$\beta = 1.15, z = 4.19, p < 1e\text{-}04$	$\beta = 0.22, z = 0.80, p = .4246$	$\beta = 0.67, z = 1.49, p = .1366$	$\beta = -0.98, z = -2.19, p = .0288$
Exclude	$\beta = 3.39, z = 8.47, p < 1e\text{-}16$	$\beta = 0.56, z = 1.72, p = .0855$	$\beta = 0.42, z = 1.31, p = .1912$	$\beta = 0.79, z = 1.42, p = .1555$	$\beta = -0.92, z = -1.66, p = .0962$

Note: The model we used is $probability \sim 1 + p(C|A) \times \Delta p3 + (1 + p(C|A) \times \Delta p3 \times group|\ participants) + (1 + p(C|A) \times \Delta p3 \times group\ |items)$. $\Delta p3$ was categorically coded as negative relevance, irrelevance, and positive relevance. When conducting the analyses, $\Delta p3$ was dummy-coded with negative relevance as the baseline. *group* signifies whether participants were required to judge the probability or acceptability of the conditional statement

(using GLMM rather than LMM): When all the data were included in the analyses, both the main effect of the parameter and its interaction with the conditional probability were observed. When the boundary cases were excluded from the analyses, the only significant effect remaining was the conditional probability $p(C|A)$.

Source of the modulation effect

We now focus on the modulation parameter $\Delta p3$. Modulation parameter $\Delta p3$ is not orthogonal to $p(C|A)$ but is a linear combination of two conditional probabilities $p(C|A)$ and $p(C|\neg A)$—that is, $\Delta p3 = p(C|A) - p(C|\neg A)$. The modulation effect of $\Delta p3$ can be reduced to $p(C|A)$, $p(C|\neg A)$, or both. If it is partially reduced to $p(C|\neg A)$, then $p(C|\neg A)$ should also have an effect. If the effect of $p(C|\neg A)$ was not observed, the effects of $\Delta p3$ should be solely reduced to that of $p(C|A)$. The interaction of $\Delta p3$ with $p(C|A)$ should actually be the high-order effects of the conditional probability $p(C|A)$—that is, $p(C|A) \times \Delta p3 = p(C|A) \times (p(C|A) + p(C|\neg A)) = p(C|A)^2 + p(C|A) \times p(C|\neg A)$. To explore this possibility, we fitted the model: $Probability \sim 1 + p(C|A) + p(C|A)^2 + p(C|\neg A) + (1 + p(C|A) + p(C|A)^2 + p(C|\neg A)\ |\ participants) + (1 + p(C|A) + p(C|A)^2 + p(C|\neg A)\ |\ images)$ to our data, either including or excluding the empty subsets. The results were summarized in Table 6. The conditional probability $p(C|A)$ had a linear-order effect regardless of whether the empty subsets were included. The second-order effect of $p(C|A)$ existed only when empty subsets were included in the analyses. The other conditional probability $p(C|\neg A)$ did not have an effect, whether the empty subsets were included or not. This is in accordance with our earlier observation that the parameter $\Delta p1$ defined from the linear combination of the four subsets did not have a modulation effect. These results suggest that the modulation effect of $\Delta p3$ actually reflects some nonlinear high-order effect introduced by the conditional probability $p(C|A)$.

Discussion

First, the modulation effects observed in intensional probability studies were successfully replicated in our study using extensional probabilities. The trick is to include boundary cases in the test stimuli (one or more empty subsets) to trigger participants' boundary responses. In intensional probability studies, no information is experimentally provided to trigger participants' responses. Participants must mentally simulate several cases and make their judgments based on the simulated scenarios. Restricted by the limited capacity of mental information processing, the simulated sample size is relatively small (Zhu et al., 2020), especially when the judged statement is complex, such as conditionals. A small sample is more probable to be a boundary one (i.e., containing one or more empty subsets). The boundariness of the simulated sample cannot be directly observed, but it is reflected in participants' boundary responses. Parallel to extensional studies,

Table 6 Effects of $p(C|A)$, $p(C|A)^2$, and $p(C|\neg A)$

Empty subsets	$p(C\|A)$	$p(C\|A)^2$	$p(C\|\neg A)$
Include	$\beta = 5.26, z = 9.20, p < 1e\text{-}19$	$\beta = -1.11, z = -2.19, p = .0284$	$\beta = 0.01, z = 0.13, p = .8966$
Exclude	$\beta = 2.69, z = 3.12, p = .0018$	$\beta = 0.19, z = 0.23, p = .8196$	$\beta = 0.01, z = 0.04, p = .9708$

removing boundary responses from analyses will eliminate the parameter's modulation effect.

Second, the modulation effect of *Δp3* actually reflects the second-order effect of the conditional probability. This modulation effect, a second-order effect of the conditional probability, can only be observed when boundary cases are included in the analysis. If we assume the true relationship between the probability of the conditional and the conditional probability is linear, the modulation effect means that the linear relationship was distorted by the occurrence of boundary cases. In our experiment, a boundary case (where one or more subsets are empty) may push the conditional probability to the boundary (0 or 1). When the conditional probability is used to predict the probability of the conditional, the probability of the conditional is then overestimated or underestimated with regard to the conditional probability. This is corroborated by the observation that judgment of probabilities tends to be distorted at or near the probability boundaries (Hilbert, 2012; i.e., people tend to underestimate high values and overestimate low ones).

Third, it is acknowledged that a sense of relation between the antecedent and the consequent is normally obtained when encountering a conditional statement. The results of the current experiment, however, remain neutral on whether the relevance effect is a kind of semantics like the conventional implicature (Douven, 2008; Douven et al., 2018; Krzyżanowska & Douven, 2018; Krzyżanowska et al., 2017; Skovgaard-Olsen, 2016), or a kind of pragmatics like the conversational implicature (Grice, 1989; Quine, 1950) or the discourse coherence (Cruz et al., 2016; Lassiter, 2022, 2024; Lassiter & Li, 2024; Over, 2023). The current experiment even remains neutral on whether the relevance effect is a necessary component or not for a conditional statement. What we observed in the current experiment is that the evidence reported in the literature to support the occurrence of the relevance effect may be confounded by other factors. These observations have two implications: First, the evidence reported to support the occurrence of the relevance effect is not solid enough. Second, cases with the same relevance effect may have different psychological statuses depending on whether the specific case falls at the boundary or not.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-025-02725-2.

Acknowledgements We would like to thank Dr. David Over and the anonymous reviewer for their insightful and detailed review. We would also like to thank Dr. David Sewell for skillfully handling our manuscript and providing useful editorial input.

Author contributions L.Z.: Funding acquisition, Project administration, Supervision, Conceptualization, Methodology, Formal analysis, Data Curation, Visualization, Writing—Original Draft, Review & Editing; M.W.: Methodology, Software, Investigation, Resources, Writing—Original Draft.

Funding This research was supported by the National Social Science Fund of China [Grant No. 19BYY087] to Likan Zhan.

Data availability The studies were not preregistered. All study materials and data have been uploaded to the Open Science Framework (OSF) and can be accessed online (https://osf.io/t5xr3/?view_only=7bbb0782f44c4865a5479cc6fac254dd).

Code availability All analysis codes are publicly available online (https://osf.io/t5xr3/?view_only=7bbb0782f44c4865a5479cc6fac254dd).

Declarations

Ethics approval This study was carried out in accordance with the recommendations of Beijing Language and Culture University Committee on Human Research Protection with written informed consent from all subjects. The protocol was approved by the Beijing Language and Culture University Committee on Human Research Protection.

Consent to participate Subjects gave written informed consent in accordance with the Declaration of Helsinki.

Consent for publication Participants gave consent for their anonymized data to be used in academic publications and to be made publicly available via OSF.

Conflicts of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Adams, E. W. (1965). The logic of conditionals. *Inquiry, 8*(1), 166–197. https://doi.org/10.1080/00201746508601430

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*(3), 147–149. https://doi.org/10.3758/bf03334492

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review, 59*(1), 65–98. https://doi.org/10.1137/141000671

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367–405. https://doi.org/10.1037/0033-295x.104.2.367

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review, 121*(3), 463–480. https://doi.org/10.1037/a0037010

Crupi, V., & Iacona, A. (2021). Three ways of being non-material. *Studia Logica, 110*(1), 47–93. https://doi.org/10.1007/s11225-021-09949-y

Douven, I. (2008). The evidential support theory of conditionals. *Synthese, 164*, 19–44. https://doi.org/10.1007/s11229-007-9214-5

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive Psychology, 101*, 50–81. https://doi.org/10.1016/j.cogpsych.2017.09.002

Evans, J. S., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 29(2), 321–335. https://doi.org/10.1037/0278-7393.29.2.321

Frege, G. (1960). Begriffsschrift. In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 1–20). Basil Blackwell.

Fugard, A. J., Pfeifer, N., Mayerhofer, B., & Kleiter, G. D. (2011). How people interpret conditionals: Shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 635–648. https://doi.org/10.1037/a0022329

Girotto, V., & Johnson-Laird, P. N. (2004). The probability of conditionals. *Psychologia, 47*(4), 207–225. https://doi.org/10.2117/psysoc.2004.207

Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin, 138*(2), 211–237. https://doi.org/10.1037/a0025940

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106*(1), 62–88. https://doi.org/10.1037/0033-295X.109.4.722

Krzyżanowska, K., Collins, P. J., & Hahn, U. (2017). Between a conditional's antecedent and its consequent: Discourse coherence vs. probabilistic relevance. *Cognition, 164*, 199–205. https://doi.org/10.1016/j.cognition.2017.03.009

Krzyżanowska, K., & Douven, I. (2018). Missing-link conditionals: Pragmatically infelicitous or semantically defective? *Intercultural Pragmatics, 15*(2), 191–211. https://doi.org/10.1515/ip-2018-0004

Lassiter, D. (2022). Decomposing relevance in conditionals. *Mind & Language. Advance Online Publication.* https://doi.org/10.1111/mila.12418

Lopez-Astorga, M., Ragni, M., & Johnson-Laird, P. N. (2021). The probability of conditionals: A review. *Psychonomic Bulletin & Review, 29*(1), 1–20. https://doi.org/10.3758/s13423-021-01938-5

Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology, 71*, 305–330. https://doi.org/10.1146/annurev-psych-010419-051132

Oberauer, K., Weidenfeld, A., & Fischer, K. (2007). What makes us believe a conditional? The roles of covariation and causality. *Thinking & Reasoning, 13*(4), 340–369. https://doi.org/10.1080/13546780601035794

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 680–693. https://doi.org/10.1037/0278-7393.29.4.680

Over, D. (2023). Independence and rationality. *The Reasoner: Focus Issue on the Handbook of Rationality, 17*(6), 47–48. https://doi.org/10.7551/mitpress/11252.001.0001

Over, D. E., & Evans, J. S. B. T. (2024). Human reasoning. *Cambridge University Press*. https://doi.org/10.1017/9781009495349

Over, D. E., Hadjichristidis, C., Evans, J. S., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology, 54*(1), 62–97. https://doi.org/10.1016/j.cogpsych.2006.05.002

Quine, W. V. O. (1950). *Methods of logic*. Holt.

Rott, H. (2019). Difference-making conditionals and the relevant Ramsey test. *The Review of Symbolic Logic, 15*(1), 133–164. https://doi.org/10.1017/s1755020319000674

Russell, B. (1906). The theory of implication. *American Journal of Mathematics, 28*(2), 159–202.

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge University Press.

Sheps, M. C. (1958). Shall we count the living or the dead? *The New England Journal of Medicine, 259*(25), 1210–1214. https://doi.org/10.1056/NEJM195812182592505

Skovgaard-Olsen, N. (2016). Motivating the relevance approach to conditionals. *Mind & Language, 31*(5), 555–579. https://doi.org/10.1111/mila.12120

Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and conditionals. *Psychological Review, 126*(5), 611–633. https://doi.org/10.1037/rev0000150

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition, 150*, 26–36. https://doi.org/10.1016/j.cognition.2015.12.017

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2017). Relevance and reason relations. *Cognitive Science, 41*(Suppl 5), 1202–1215. https://doi.org/10.1111/cogs.12462

Stalnaker, R. C. (1970). Probability and conditionals. *Philosophy of Science, 37*(1), 64–80.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315. https://doi.org/10.1037/0033-295x.90.4.293

van Rooij, R., & Schulz, K. (2018). Conditionals, causality and conditional probability. *Journal of Logic, Language and Information, 28*(1), 55–71. https://doi.org/10.1007/s10849-018-9275-5

White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 710–727. https://doi.org/10.1037/0278-7393.29.4.710

Zhu, J. Q., Sanborn, A. N., & Chater, N. (2020). The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review, 127*(5), 719–748. https://doi.org/10.1037/rev0000190

Alday, P., & Bates, D. (2025). *MixedModels.jl*: A Julia package for fitting (statistical) mixed-effects models. *Zenodo*. https://doi.org/10.5281/zenodo.15053987

Cruz, N., Over, D., Oaksford, M., & Baratgin, J. (2016). Centering and the meaning of conditionals. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society. Cognitive Science Society* (Vol. 38, pp. 1104–1109). eScholarship University of California.

Douven, I. (2015). *The epistemology of indicative conditionals: Formal and empirical approaches*. Cambridge University Press.

Douven, I., Elqayam, S., & Mirabile, P. (2022). Inference strength predicts the probability of conditionals better than conditional probability does. *Journal of Memory and Language*, *123*. https://doi.org/10.1016/j.jml.2021.104302

Edgington, D. (1995). On conditionals. *Mind: A Quarterly Review of Philosophy*, *104*(414), 235–329. https://doi.org/10.1093/mind/104.414.235

Jeffrey, R. (2006). *Formal logic: Its scope and limits* (4 ed.). Hackett Publishing Company.

Lassiter, D. (2024). Distinguishing semantics, pragmatics, and reasoning in the theory of conditionals. *Inquiry*, 1–22. https://doi.org/10.1080/0020174x.2024.2405186

Lassiter, D., & Li, C. (2024). Pragmatic intrusion in probability judgment: The case of conditionals. In L. K. Samuelson, S. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society* (Vol. 46, pp. 1464–1471). eScholarship University of California.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51,* 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rott, H. (1986). Ifs, though, and because. *Erkenntnis*, *25*(3). https://doi.org/10.1007/bf00175348

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

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford University Press.