How generics obscure the logic of conditionals

Abstract

Indicative conditionals containing modal operators and adverbs of quantification generate apparent counter-examples to core logical principles including modus ponens and modus tollens. We use results from natural language semantics to show that the counter-examples are fallacies of equivocation, generated by a semantic interaction between conditional antecedents and other operators. Conditionals with a generic or habitual interpretation generate similar spurious counter-examples. As a result, results that appear to be informative about the logic of conditionals may actually reflect an interaction between conditionals and a covert generic operator.

As a second illustration of the importance of controlling for generic and habitual interpretation in the study of conditionals, we consider a recent controversy in conditional semantics in which a prominent theory of conditionals due to de Finetti appears to make plainly incorrect predictions in certain cases. In an experimental study, we show that the problematic cases arise only when the stimuli receive a generic/habitual interpretation. The predictions of de Finetti's theory are in fact correct when we restrict attention to episodic conditionals, and it is readily extended to give a compositional explanation of why superficially similar arguments involving generics behave differently. These results underline the importance of careful attention to compositional semantics when constructing and analyzing reasoning experiments.

1. Apparent counter-examples to core principles of conditional logic

A large body of psychological research has investigated the status of logical principles such as modus ponens (MP) and modus tollens (MT).

(1) MP: If p then q. p. Therefore, q.

(2) MT: If p then q. Not q. Therefore, not p.

The general finding is that experimental participants reliably endorse instances of MP in a wide range of experimental settings. Endorsement rates for MT are generally lower, in the 70%-range (see, e.g., Evans and Over, 2004 for a survey). Since most theories of conditionals validate both MP and MT, much theoretical effort has been devoted to explaining in psychological terms why instances of MT inferences are not as robustly endorsed as MP.

On the other hand, one unambiguous counter-example is enough to show that an argument pattern is not valid. Here is an apparent counter-example to MP:

⋆

[1]

(3) a. If it's raining, Mary usually takes the bus to work.
b. It's raining.
c. Therefore, Mary usually takes the bus to work.

This argument fits the MP template, but its premises clearly do not warrant its conclusion. Suppose that Mary lives in a place where it rarely rains. She never take the bus in the typical non-rainy weather, but she generally does so in the occasional instances of rainy weather. Right now, it happens to be raining. Then (3a) and (3b) are both true, but (3c) is false: usually, it is not raining, and Mary does not take the bus.

A similar argument appears to establish the invalidity of MT (see Cantwell, 2008a; Yalcin, 2012).

(4) a. If it's raining, Mary usually take the bus to work.
b. Mary does not usually take the bus to work.
c. Therefore, it's not raining.

Plainly, the truth of the premises of (4) tells us nothing about the current state of the weather.

It would be a major coup if the simple arguments in (3) and (4) could establish that MP and MT are not valid, contrary to the findings of decades of logical and psychological research. Fortunately, this conclusion is not warranted. The first part of this paper explains why, using results from natural language semantics. The second part shows that the same problems arise in the study of generic language, and that this fact presents additional methodological difficulties because generics are not overtly marked in English and many other languages. As a result, failure to control for generic interpretation can lead theorists to reject logical principles that are in fact valid, based on merely apparent counter-examples.

Section 3 demonstrates this point in a case study involving generic and habitual conditionals that appear at first glance to refute modus ponens and modus tollens. In a second case study, section 4 considers a somewhat more complex case involving the interpretation of "nested" conditionals—those with other conditionals in the antecedent and/or consequent. An otherwise attractive proposal by de Finetti (1936) makes predictions about logical equivalencies among different types of conditionals that seem at first glance to be obviously incorrect. However, the obvious counter-examples uniformly involve generic and habitual conditionals. We present the results of an experiment showing that the equivalencies predicted by de Finetti are in fact endorsed by participants in single-case, non-generic nested conditionals. This result suggests that the problematic inference patterns are in fact valid, and the appearance of invalidity is generated by a semantic interaction between the *if*-clause and other operators.

Both case studies illustrate a larger point: careful attention to semantic interactions between conditionals and generic/habitual language is crucial in understanding the basic logic of conditionals. Conversely, failure to attend to this subtle distinction in linguistic/philosophical argumentation and in the design of experimental materials can lead to incorrect logical conclusions and confounded experimental data. Theorists should be extremely

cautious in the use of generic language in work on reasoning, because generic examples and stimuli can obscure the logic of conditionals.

2. Diagnosing the counter-examples: Semantic background

Philosophers and linguists working on natural language semantics have paid close attention to semantic interactions between conditionals and other operators, such as modals (e.g., *must, possibly*) and quantificational adverbs (e.g., *usually* in (3) and (4)). Lewis (1975) showed that the conditional antecedent in examples like (5) serves to restrict the domain over which the underlined operator quantifies.

(5) If it rains, I usually take the bus to work.

(5) expresses the claim that most of the situations in which it rains are ones in which the speaker takes the bus. It is not about what usually happens *in general*, and in particular it makes no claim about what the speaker does when it is not raining.

Lewis (1975) explained this interaction using the tools of compositional semantics. Adverbs like *usually, never*, and *always* express quantifiers over relevant situations: for instance, *Mary usually drinks beer* means that most of the relevant situations are ones in which Mary drinks beer (e.g., von Fintel, 2004). Lewis showed that the *if*-clause in (5) restricts the domain of the *most*-quantifier expressed by *usually*. If R is the set of contextually relevant situations for the interpretation of (5), then *usually* in (5) quantifies over the restricted set $R \cap \textbf{rain}$: the relevant situations in which it is raining. The compositional interpretation of (5) is therefore that most situations in which it rains are ones in which Mary takes the bus.

In formal notation, (5) has the semantic analysis in (6), and the *if*-clause is analogous to the underlined portion of the quantificational sentence in (7a). (The variable D in (7b) refers to the quantificational domain of "most", i.e., the objects that are currently relevant for the purposes of the conversation.)

(6) MOST [$s \in R$, it rains in s] [I take the bus in s]

(7) a. Most dogs bark.

b. MOST [$x \in D$, x is a dog] [x barks]

Just as (7a) says nothing about non-dogs, the literal interpretation of (5) says nothing about non-rain situations (though it may well carry certain implicatures).

Kratzer (1991) shows that Lewis' observation extends also to modal sentences like (8), which expresses that the probability of regret is high *on the assumption* that Mary purchases a ticket.

(8) If Mary buys a lottery ticket, she'll probably regret it.

(9) But she probably won't regret anything, since she rarely buys lottery tickets.

(8) is compatible with the continuation in (9), which would not be possible if the two instances of *probably* were talking about the same body of information. The instance of *probably* in (9) is about what is likely given everything we know, while the instance in (8)—restricted by

an *if*-clause—is talking about what's likely given what we know, plus the assumption that Mary buys a ticket.

While this much is well-established in the literature on natural language semantics, it has received virtually no attention in the voluminous psychological literature on conditionals. In this paper, we will show that semantic interactions of this kind are theoretically and methodologically important not only for linguistics and philosophical logic, but also for the psychological study of conditionals.

The crux of the problem is that the domains of modals and quantificational adverbs—*probably*, *must*, *usually*, and *always*, for instance—are determined by both context and semantic interactions within a sentence. Assessing argument validity requires holding quantificational domains constant, along with other context-sensitive aspects of interpretation (e.g., pronouns, indexicals, and tense, as well as lexical and structural ambiguities). If an expression occurs multiple times in an argument, but its context-sensitivity is resolved in different ways in different instances, the argument may be a fallacy of equivocation. For example, the argument in (10) has the surface form of MP, but it is not a compelling argument unless both instances of the pronoun *she* are resolved as referring to the same individual.

(10) a. If it's late, she is tired.
b. It's late.
c. Therefore, she is tired.

If *she* refers to Hillary Clinton in (10a) and to Angela Merkel in (10c), then the premises could easily be true while the conclusion is false. This is not, of course, a reason to reject MP. (10) is only a genuine MP argument when the two instances of *she* are resolved in the same way.

Standardly, linguists and philosophers deal with this issue by appealing to some notion of "logical form" which resolves ambiguities and context-dependent aspects of interpretation (Harman, 1972; Stanley, 2000). Whether or not such logical forms play a role in the process of human reasoning (see Braine, 1978; Braine et al., 1984 and Johnson-Laird, 2010 for contrasting positions), some such notion is needed in determining whether an argument like (10) even counts as an instance of MP. The reason is that a notion of argument validity based on the uninterpreted form of strings would make *every* argument form invalid. As (10) shows, MP is readily falsified on this interpretation. Even the principle of Identity (from p, infer p) is easily falsified on a string-based picture of reasoning: from *He is 57 years old*, uttered at some time while pointing to one person, we cannot infer that *He is 57 years old* is true when uttered at a different time, or while pointing to a different person.

In order to adjudicate a purported counter-example to an argument form, then, we must first ensure that we are working with a formal representation of the sentence in which all shiftable and context-dependent elements of interpretation have been resolved: indexical reference, lexical ambiguities, and operator domains, among others. This intermediate level of representation, syntactic in character but with context-dependent elements of meaning resolved, is what we consider here as "Logical Form".

This point is important in the analysis of the purported counter-examples to MP and MT given above, because the domain variables are not held constant in the logical forms of

these arguments. Logical forms of quantified sentences must include a specification of the domains of operators like *usually* (Stanley, 2000; Stanley & Szabó, 2000). The analysis of Lewis (1975) implies that the two instances of *usually* in arguments (3) and (4) are associated with different domains. As a result, these arguments are fallacies of equivocation, like an interpretation of (10) where the instances of *she* are resolved in different ways. The only difference is that the domain shift is due to a compositional interaction between elements of the sentence, rather than by a shifting context.

More precisely, let s be a variable over situations, and R represent a restriction to contextually relevant situations. Then our purported counter-example to MP in (11) has the logical form in (12).

(11) a. If it's raining, Mary usually takes the bus to work.
b. It's raining.
c. Therefore, Mary usually takes the bus to work.

(12) a. MOST $[s \in R$: it's raining in $s]$ [Mary takes the bus in $s]$
b. It is raining (at the time of utterance)
c. MOST $[s \in R]$ [Mary takes the bus in $s]$

It is clear why this is not a valid argument. (12a) claims that most relevant situations *in which it is raining* are ones in which Mary takes the bus. In contrast, the proposed conclusion (12c) claims that most relevant situations are ones in which she takes the bus, no matter the weather. The instances of *usually* in the major premise and in the conclusion have different interpretations because the former interacts semantically with the antecedent of the conditional.[1] Lewis' analysis of adverbs like *usually* makes it clear why (3) is invalid: this apparent counter-example to MP is a fallacy of equivocation.

A similar analysis accounts for the apparent counter-example to MT in (4): the two instances of *usually* have different domains. The instance in the major premise (4a) is restricted by the conditional antecedent *it's raining*, and so quantifies over contextually relevant rain situations. The instance of *usually* in the minor premise (4b), in contrast, quantifies over all contextually relevant situations. Because the domains are resolved differently, (4) is not a genuine instance of MT. Here again, an apparent counter-example to a core logical principle involving conditionals is a fallacy of equivocation generated by the semantic interaction of conditional antecedents with other operators. As Cantwell (2008a) discusses, similar apparent counter-examples to MT are readily found involving deontic and ability modals, with the same diagnosis.

Summing up so far, conditionals interact semantically with operators that express generalizations and various types of modality. Making the logical forms of arguments like (3) and (4) explicit removes the temptation to treat them as invalidating core principles of conditional logic. This point is also methodologically important for research on reasoning since

[1] To make matters worse, the instance of the string *It is raining* also has a context-dependent interpretation, picking out a class of events in (3a) but a single event in (12b) that is spatiotemporally anchored to the speech event.

such compositional interactions can introduce confounds in the materials designed to evaluate the status of logical principles. As we will see in the next section, sentences that express generalizations with no overt linguistic marking are particularly problematic in this respect.

3. Case Study 1: Reasoning with generics and habitual conditionals

Generic and habitual sentences express generalizations about the properties of objects or classes of repeated events. They are semantically related to sentences containing adverbs of quantification but differ in that they are not overtly marked as conveying generalizations in English and many other languages (Carlson, 2012; Krifka et al., 1995). For instance:

(13) Generics:
   a. Monkeys like bananas.
   b. Major European cities have cathedrals.

(14) Habituals:
   a. I take the bus to work.
   b. Mary eats broccoli.

Roughly, generics are about typical properties of kinds of individuals, while habituals are about individual dispositions or properties of kinds of events. Crucially, nothing in the overt form of these sentences explicitly indicates this aspect of their meaning.[2]

Generic sentences have been of considerable interest to philosophers, linguists, and computer scientists for a variety of reasons (Carlson & Pelletier, 1995). They have also received a great deal of attention in recent psychological work (Berio & Musholt, 2023; Gelman, 2021; Leslie, 2017). In natural language semantics, generic sentences are typically analyzed as containing a covert adverbial quantifier GEN, which has an interpretation along the lines of "generally" or "usually". On this analysis, the logical form of the sentences in (13) and (14) is along the following lines (with R once again representing a pragmatically supplied class of relevant situations).

(15) a. (13a) $\rightsquigarrow$ GEN [x is a monkey] [x likes bananas]
   "Generally, if something is a monkey, it likes bananas"
   b. (14a) $\rightsquigarrow$ GEN [$s \in R$] [I take the bus to work in s]
   "Generally, among relevant situations (e.g., those in which I need to go to work), I take the bus to work"

[2]Many languages, including English, do have means of overtly marking of generic and/or habitual marking: for instance, the English *used to* construction (Carlson, 2012; Comrie, 1976) or the Shiiluk verbal suffix *-nyi* (Remijsen et al., 2024). In English, the use of the simple present often favours a generic/habitual interpretation. This is in contrast to many other languages, for instance, French and German, in which the simple present can refer to an ongoing event or one in the near future and listeners use context to disambiguate generic vs. episodic interpretation.

Notably, generics leave room for counter-instances. For example, (13a) contrasts with *All monkeys like bananas* in that the generic sentence can be true even if there are a few monkeys who do not like bananas.

On this prominent semantic analysis, then, generic interpretation is due to the presence of a covert adverb of quantification in the logical forms of the relevant sentences. If so, we would expect that GEN should show a semantic interaction with conditionals, as adverbs of quantification do. This is correct: as a number of authors have noted, conditional antecedents can restrict the domain of GEN (Farkas & Sugioka, 1983; Krifka et al., 1995; Lassiter & Baratgin, 2021).

(16) a. If it rains, I take the bus to work.
b. "Generally, in relevant situations *in which it rains*, I take the bus."
c. GEN [$s \in R$, **it rains in** s] [I take the bus in s]

(17) a. Bartenders are happy if they get big tips.
b. "Generally, in relevant situations *in which a bartender gets a big tip*, the bartender is happy."
c. GEN [$s \in R$, bartender x, x **gets a big tip in** s] [x happy in s]

In the schematic logical forms (16c) and (17c), the boldfaced material corresponds to the antecedent of the conditional. This material functions to restrict the domain of situations to which the generalization applies. This interaction between GEN and conditional antecedents is semantically identical to the interaction with modals and adverbs of quantification discussed in the previous section.

The connection between GEN and adverbs of quantification suggests that we may find spurious counter-examples to core logical principles in generic conditionals as well. Indeed we do: for example, here is a spurious counter-example to MT that relies on a generic conditional.

(18) a. Mary doesn't take the bus to work—she cycles.
b. If it's raining very hard, Mary takes the bus to work.
c. Therefore, it's not raining very hard.

(18a) and (18b) are consistent thanks to the exception-tolerance of generics and habituals: both can be true of a single person. And both of these generalizations can hold together without implying anything at all about the current state of the weather, including the candidate MT inference in (18c). The explanation of this pattern is now familiar. (18a) contains a GEN quantifier that ranges over all relevant situations, while (18b) contains an instance of GEN whose domain is restricted to situations in which it rains very hard. (18c), on the other hand, does not contain GEN at all: it is a single-case, "episodic" statement making a claim about the state of weather at the time of utterance.

Similarly, (19) provides a spurious counter-example to MP involving a conditional with a habitual interpretation.

(19) a. Johnny eats pizza.

b. If Johnny eats pizza, he feels happy.

c. Therefore, Johnny feels happy.

(19) describes a habitual property of Johnny: he is disposed to consume pizza in the right circumstances (e.g., when it is available, he has not already had dinner, he is not sick, etc.). (19b) describes how Johnny generally feels in situations in which he eats pizza: the generalization is restricted to situations that satisfy the antecedent. Neither of these generalizations license any inferences about how Johnny feels right now, as (19c) would have it: we would need to know something about what Johnny has recently had to eat, and this information is not present in the premises. Here again, the compositional analysis sketched above clarifies why this apparent instance of MP does not constitute a genuine counter-example. The domain of the generic/habitual operator is different in (19a) and (19b), and no such operator is even present in (19c), which is oriented toward the utterance time.

The failure of the arguments in (18) and (19) further underlines the necessity, noted above, of attending to logical form when evaluating proposed instances of a logical principle. Since a single counter-example is enough to show that an argument form is not valid, the clear invalidity of (18) and (19) would show that MP and MT are not valid principles of reasoning—if these were genuine instances of MP and MT. As we argued for similar arguments with modals and adverbs above, these are spurious arguments because the instances of GEN in the premises are associated with different quantificational domains. If we were to focus on the overt form of the sentences without attending carefully to the semantic interpretation, we might miss this subtle point and draw the wrong conclusion about the status of a core principle of conditional logic.

More generally, since GEN is not overtly realized in English, many sentences are ambiguous between a generic interpretation and a single-case episodic interpretation. Linguistic tests—some of which are discussed below—can help us to discern whether GEN is present in the logical form of a given sentence. Fixing the logical form of each sentence in an argument is a prerequisite for knowing whether the argument counts as an instance of MP, MT, or some other argument schema. On top of the interaction with conditionals that we have focused on, there is also the very general problem of checking that each two instances of a string in an argument are matched in the presence or absence of the GEN operator in their respective logical forms.

The interaction of conditionals and generic interpretation leads to two important conclusions. Theoretically, we have reason to continue treating MP and MT as valid argument forms, despite a second round of apparent counter-examples. Methodologically, the examples reveal an important confound that has not been carefully controlled in the vast philosophical and psychological literature on conditionals. Because of the semantic interaction of conditionals and GEN, we cannot draw conclusions about the basic logic of conditionals from examples involving generic or habitual sentences, or from experiments that use generic and habitual sentences as stimuli. In order to avoid confounds when testing theories about argument validity, we must ensure that the stimuli are uniformly non-generic.

4. Case study 2: Objections to the de Finetti semantics

In this section, we introduce a more complex example in which attention to compositional semantics and the generic/episodic distinction is crucial to evaluating empirical arguments about the validity of argument forms. The focus is a recent controversy about certain semantic equivalences that are predicted by the three-valued conditional semantics of de Finetti (1936). The theoretical and methodological issues that arise are strikingly similar to the ones that came up in our discussion of MP/MT inferences in previous sections. In both cases, a theory of conditionals that has enjoyed much empirical support is threatened by what appear to be compelling counter-examples. As the experiment reported below will show, the problematic inference patterns in question are indeed rejected when generic sentences are used. However, sentences with the same overt form are accepted when we implement linguistic manipulations that enforce a single-case, episodic interpretation. This turns out to be exactly what is predicted by the de Finetti semantics, together with theories of genericity and domain restriction from modern linguistics. As a result, the contested inference patterns are on sound footing, and the appearance of counter-examples is an artifact of the semantic interaction of conditionals and genericity.

4.1. The three-valued de Finetti semantics

The three-valued truth-functional semantics proposed by de Finetti (1936)[3] holds that conditionals with true antecedents have the truth-values of their consequents; otherwise, they are undefined or "void". That is, *If A, C* is

- true if A and C are both true;
- false if A is true and C is false;
- otherwise undefined.

In de Finetti's proposal, a key concern was to find a way of validating the equation between the probability of a conditional "If A, C" and the conditional probability $P(C \mid A)$. This equation has much intuitive appeal and was subsequently the subject of a large philosophical literature (Adams, 1975; Edgington, 1995; Lewis, 1976). It is by now well-supported empirically (Douven & Verbrugge, 2010; Over & Evans, 2003), and the de Finetti semantics has the virtue of validating it straightforwardly (Cantwell, 2008b; Lassiter, 2020; Milne, 1997; Mura, 2009).

The trivalent semantics for conditionals was rediscovered numerous times in psychology and philosophical logic in the decades following de Finetti's proposal. In psychological work, it was proposed by Wason (1966) as the "defective truth-table" (see Over and Baratgin 2017; Over and Evans 2024). Wason adopted the third value in response to experimental participants' tendency to judge false-antecedent cases as irrelevant to the truth-value of a conditional, rather than true or false. In recent years there has been a great deal of

[3]The theory was formulated in detail as early as 1928 by de Finetti: see Baratgin, 2021.

formal and empirical work on trivalent logics for conditionals. Logicians have explored de Finetti's theory and its variants (Egré & Rossi, 2024; Égré et al., 2021; Joaquin, 2023; Milne, 1997; Mura, 2009, 2011, 2021; Vidal, 2014), and psychologists have shown that the trivalent semantics can account for a variety of experimental data (Baratgin et al., 2013, 2014; Baratgin et al., 2018; Evans & Over, 2004; Politzer et al., 2020; Politzer et al., 2010). Linguists and philosophers have also explored the use of three-valued semantics to model the use of conditionals to restrict quantifier domains (Belnap, 1970; Huitink, 2008; Lassiter & Baratgin, 2021).

While the de Finetti semantics is attractively simple and enjoys a good deal of empirical support, Douven (2016) points out a number of reasons to be concerned about its correctness. Douven shows that this semantics validates several argument forms that seem to admit of obvious counter-examples. Specifically, he proves that de Finetti's semantics renders conditionals of the following three forms logically equivalent.[4]

(20) a. If (A and B), then C. [conjunctive]
b. If A, then (if B then C). [right-nested]
c. If (B if A), then C. [left-nested]

The equivalence between (20a) and (20b) is known in the literature as the "Import-Export" property and is widely thought to be empirically correct. However, Douven (2016) presents examples that call into question the equivalence of left-nested conditionals (20c) with either conjunctive or right-nested examples. For example, the sentences in the following pair are clearly not semantically equivalent.

(21) a. If this material gets hot and becomes soft, it is not suitable for our purposes.
b. If this material becomes soft if it gets hot, it is not suitable for our purposes.

Intuitively, (21a) is a claim about what happens if a particular material *actually* becomes hot and soft. In contrast, (21b) seems to be a claim about what happens if the material has a certain *disposition*: to become soft whenever it gets hot. Douven also shows that the de Finetti semantics predicts doubly-nested conditionals of the form "If (B if A), then (D if C)" to be equivalent to conditionals with a three-conjunct antecedent: "If (A and B and C), then D". He provides the following counter-example:

(22) a. If your mother gets angry if you come home with a B, then she'll get furious if you come home with a C.
b. If you come home with a B and your mother gets angry and you come home with a C, then your mother will get furious.

Again, the sentences in (22) are plainly not equivalent. (22a) is intuitively about the addressee's mother's dispositions—how she would *tend to* behave in various possible circumstances—

[4]The equivalence of these sentence types was in fact noted already by de Finetti in 1928 (see Baratgin, 2021) and repeated in de Finetti (1970, p.710). See de Finetti, 1975, pp.328 for an English translation.

while (22b) is about what will happen if certain concrete possibilities obtain (a B on one assessment, an angry outburst, and a C on another assessment).[5] Douven (2016) concludes:

> The predictions that de Finetti's semantics makes regarding the truth conditions and probabilities of nested conditionals deviate so starkly from what common sense suggests that one need not even conduct an experiment to know that the semantics' predictive accuracy vanishes when we turn to nested conditionals.

While the argument is valid, Lassiter and Baratgin (2021) show that it is unsound. The de Finetti semantics does predict that the sentence types in (20) are logically equivalent. However, they argue that the purported counter-examples are not in fact instances of the relevant schemata, because—in each case—one or both of the examples involve one or more layers of generic interpretation. To see this, note that the intuitive readings of the nested-conditional sentences (21b) and (22a) involve *dispositions*—generalizations about how a person or material would behave in a specified class of situations. In contrast, the sentences with conjunctive antecedents, (21a) and (22b), express claims about specific events occurring at particular times.

Specifically, for the examples in (21), Lassiter and Baratgin (2021) claim that the most natural interpretations of the two conditionals are due to the following logical forms, which are mismatched in terms of the presence or absence of a generic operator in the antecedent.

(23) a. If (this material gets hot and becomes soft), it is not suitable for our purposes.
b. If GEN(this material becomes soft if it gets hot), it is not suitable for our purposes.

The logical forms in (23) predict the intuitive difference in interpretation: (23a) is read as talking about what happens if certain events actually transpire, while (23b) is about what happens if the material has a certain *dispositional* property. The examples in (22) are similarly mismatched, with two generic operators in the first example and none in the second.

(24) a. If GEN(your mother gets angry if you come home with a B), then GEN(she'll get furious if you come home with a C).
"If your mother has the disposition to get angry in B-situations, then she has the disposition to get furious in C-situations."
b. If (you come home with a B and your mother gets angry and you come home with a C), then your mother will get furious.
"If the following things actually happen—you get a B, she gets angry, and then you get a C—then she will get furious."

[5] This interpretation of Douven's example differs from his own: he concludes, uncharitably, that (22b) has a contradictory antecedent. This seems to be due to a failure of pragmatic reasoning. In ordinary language use, the B and C grades in question could readily be associated with different assessments, making the example non-trivial.

In his writings, de Finetti insists that his semantics is only intended to apply to "specific" conditionals (de Finetti, 1936, 1979a, 1979b). In terms of modern linguistic theory, de Finetti's claims are intended to apply at least to "episodic" conditionals, which make a claim about a specific, spatiotemporally bounded event. The theory was never intended to apply to generalizations. The semantic theories of genericity discussed above lend support to de Finetti's restriction, albeit for very different reasons from his own. The semantic interpretation of a generic conditional involves quantification over a class of events, just as an overt adverbial quantifier like "generally" or "usually" would. In contrast, overtly similar episodic conditionals talk about the properties of unique events. In addition, semantic interactions between conditional antecedents and GEN noted above lead to additional semantic differences between examples with and without GEN.

If the intuitive logical forms argued for by Lassiter and Baratgin (2021) are correct, then, de Finetti's semantics does not in fact predict that the examples in (21) and (22) should be equivalent. However, their argumentation remains at the level of intuition. It also falls short of showing that de Finetti's predictions about nested conditionals are actually correct. Instead, Lassiter & Baratgin's arguments show only that the de Finetti theory is not obviously falsified by the apparent counter-examples that Douven notes.

4.2. Logical predictions

The de Finetti semantics makes the following basic predictions about logical relations among episodic conditionals:

- For episodic (single-event) conditionals, all of the following should be equivalent (Douven, 2016):
  - **Conjunctive**: If A and B, then C.
  - **Right-nested**: If A, then if B then C.
  - **Left-nested**: If B if A, then C.
- For episodic conditionals, the following should be equivalent (Douven, 2016):
  - **Doubly nested**: "If B if A, then D if C."
  - **Three-conjunct**: "If A and B and C, then D."

For generic conditionals, the de Finetti theory makes a different set of predictions when combined with theories of generic interpretation from modern linguistics (see Huitink, 2008; Lassiter & Baratgin, 2021).

- Right-nested generic conditionals with a single wide-scope GEN operator should be equivalent to conjunctive conditionals with a single wide-scope GEN operator.
  - **Conjunctive generic**: GEN(If A and B, then C).
  - **Right-nested generic**: GEN(If A, then if B then C.)

- Left-nested generic conditionals should not be systematically equivalent to conjunctive generics, or to any of the other conditional types in question.
  - **Left-nested generic**: If GEN(B if A), then C.
- Doubly-nested generic conditionals should not be systematically equivalent to three-conjunct generic conditionals.
  - **Doubly-nested**: If GEN(B if A), then GEN(D if C).

For generics, the only predicted equivalence is between right-nested and conjunctive generic conditionals. The reason for this equivalence is that these conditionals end up restricting the domain of the generic operator in the same way. Formally, this is because domain restriction is a set-intersection operation, and set-intersection is associative. In addition, conjunction is standardly modelled as a set intersection operation. As a result, if the domain of GEN is R, then successive restriction by some sets A and B is equivalent to restriction by the conjunction (intersection) of A and B:

$$(R \cap A) \cap B = R \cap (A \cap B)$$

As a result, when GEN is restricted by *if* A and then by *if* B—as in a right-nested generic conditional—its domain ends up being $(R \cap A) \cap B$. This is the same as the result if GEN's domain is restricted by $A \cap B$, as in a conjunctive generic conditional. No similar set-theoretic fact predicts the equivalence of any of the other pairs of generic sentence types under investigation. As a result, where generics are concerned, the theory predicts the equivalence of right-nested and conjunctive conditionals only.

4.3. Experimental concept

To test these predictions directly, we introduced an experimental manipulation that allowed us to compare pairs of episodic conditionals to generic pairs that were matched in relevant respects. We presented participants with a pair of conditionals, one nested and one conjunctive, and asked them to judge to what extent the conditionals had the same interpretation.

Since the predictions depend on whether the sentences have a generic interpretation, it was crucial in our experiment to control for the presence or absence of a GEN operator. A simple method of blocking a generic interpretation is to include specific time adverbials, such as "right now", "Tuesday at 3 PM", or "on January 14". Such adverbial are usually semantically incompatible with a generic interpretation. For instance, (25a) is most naturally read with a generic interpretation, as describing a habitual action; while (25b) can only be read as describing a unique, spatiotemporally located event.

(25) a. As a teenager, Bill got drunk at parties. (habitual)
b. As a teenager, Bill got drunk at his seventeenth birthday party. (episodic)

The shift from generic/habitual interpretation in (25a) to episodic interpretation in (25b) is also supported by the shift from the indefinite plural "at parties" to the singular definite "at his seventeenth birthday party". Generic interpretations are frequently less available when definite temporal or spatial reference is indicated. The use of the past tense also tends to favor an episodic interpretation, though a generic or habitual interpretation is still possible in past tense sentences (e.g., (25a)).

Our experimental design exploited these linguistic manipulations as follows. Pairs of intuitively generic conditionals were generated, without any of the specific features just noted, where one conditional had a conjunctive antecedent and the other was nested (left, right, or doubly: see predictions above). To generate matched stimuli for the episodic condition, we modified the examples to add a specific time adverbial. In some cases, we adjusted other linguistic features described above or modified other details to maintain overall coherence. Crucially, the gross form of the sentences was not changed. As a result, our experiments made it possible to test the hypothesis that episodic conditionals follow the de Finetti predictions, while pairs of generic conditionals do not.

For example, one of our scenarios collected equivalence judgments for the conjunctive and left-nested generic conditionals in (26) (= Douven's examples (21) above) and for the matched episodic pair in (27).

(26) Generic left-nested vs. conjunctive

a. If this material becomes soft if it gets hot, it is not suitable for our purposes.

b. If this material gets hot and becomes soft, it is not suitable for our purposes.

(27) Episodic left-nested vs. conjunctive

a. If this material gets soft right now if it got hot during the morning, then our workers won't be able to use it this afternoon.

b. If this material got hot during the morning and it gets soft right now, our worker won't be able to use it this afternoon.

According to the predictions described above, participants should judge that the pair of conditionals in (27) were highly similar in meaning and that the pair in (26) were less so.

4.4. Experiment

4.4.1. Participants

We collected data from 100 English native speakers (75 women, and 25 men). The average age was 37.6 years (SD: 11.4, minimum 20, maximum 78). 82 of the participants had a higher education, 16 had a high school degree and 2 did not have any degree. The participants were recruited on Prolific.com and were paid GBP £1.05 for completing the questionnaire. Ethical approval was not required for this type of experiment according to local regulations.

4.4.2. Material and procedure

The questionnaire was designed on Qualtrics. The materials consisted of 6 scenarios, with two testing each of the three types of nested conditionals (left-, right-, and doubly-nested). Each scenario consisted of a short context sentence, followed by a short exchange between two characters. The first character asked a question, and the second character always responded with a nested conditional sentence. We then present the participant with a conjunctive, (non-nested) conditional and ask the participants to evaluate how similar the meanings of the two sentences were using a slider with endpoints labelled "Absolutely different" on the left side and "Absolutely the same" on the right. Participants did not see any numerical values and had to make their choice purely on the position of the mark on the slider (see Figure 1). Responses were recorded as a proportion of the scale, with the left side scored as 0 and the right side as 100.

Each scenario had both a generic and an episodic version in which only the response of the second character (the conditional sentence) differed. As there were two scenarios per type of nesting, one was presented in its generic form and the other one in its episodic form. As a result, participants did not receive the same scenario in both generic and episodic form but still saw stimuli from all conditions. Generic and episodic forms of each scenario were distributed in two questionnaires. Participants were randomly assigned to one of them. Each item was presented on a separate page in random order.

Finally, one page was devoted to demographic questions (gender, age and level of education).

4.4.3. Data analysis

A summary of the raw results is presented in Table 1 and Fig. 2.

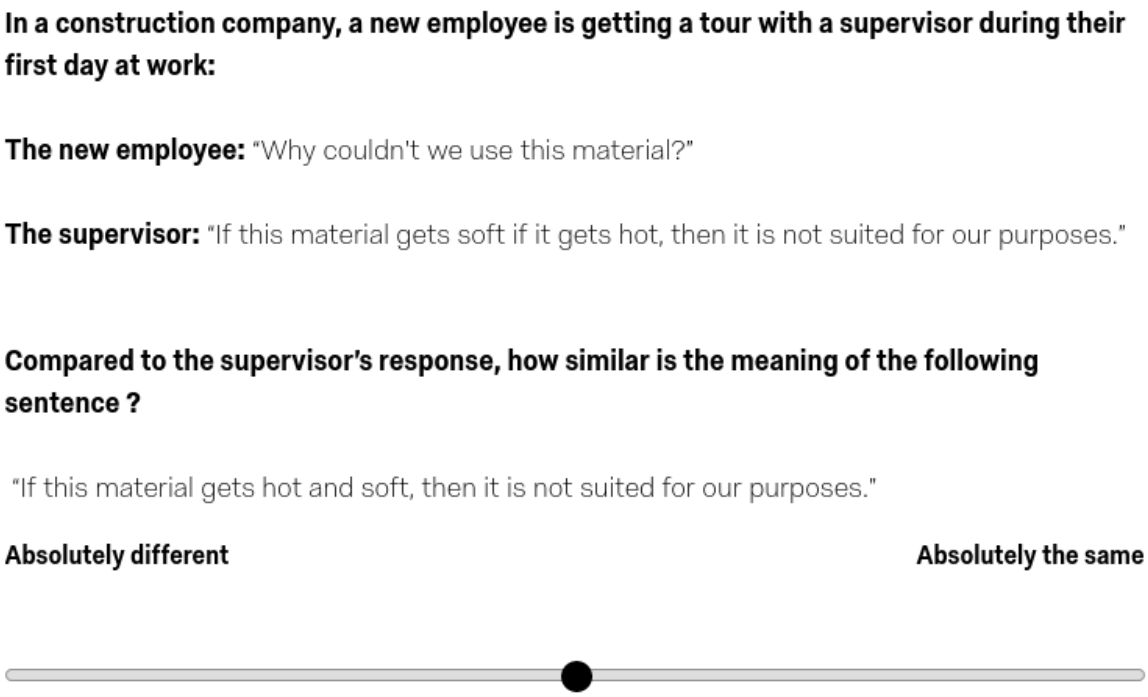
Table 1
Raw measures of the similarity between nested and simplified statements (N=600).

Event	Nesting	Mean Similarity	SD	Count
Generic	Both	55	34	100
Generic	Left	69	31	100
Generic	Right	88	20	100
Episodic	Both	80	27	100
Episodic	Left	83	22	100
Episodic	Right	83	26	100

For modelling purposes, we scaled similarity ratings to (0,1) to model them using beta distributions ($y' = y/100$) and rescaled the result to avoid bounds: $y'' = (y'(N-1) + 1/2)/N$ (Smithson & Verkuilen, 2006). 5 different mixed models were fitted using glmmTMB, all with random effects of participant and questionnaire (see Table 2): Model 0, with no predictor; Model 1, whose only predictor was the type of nesting (right-, left-, or doubly-nested) compared to the conjunctive conditional; Model 2, with only the type of event

Figure 1

Sample item: Generic left-nested vs. conjunctive



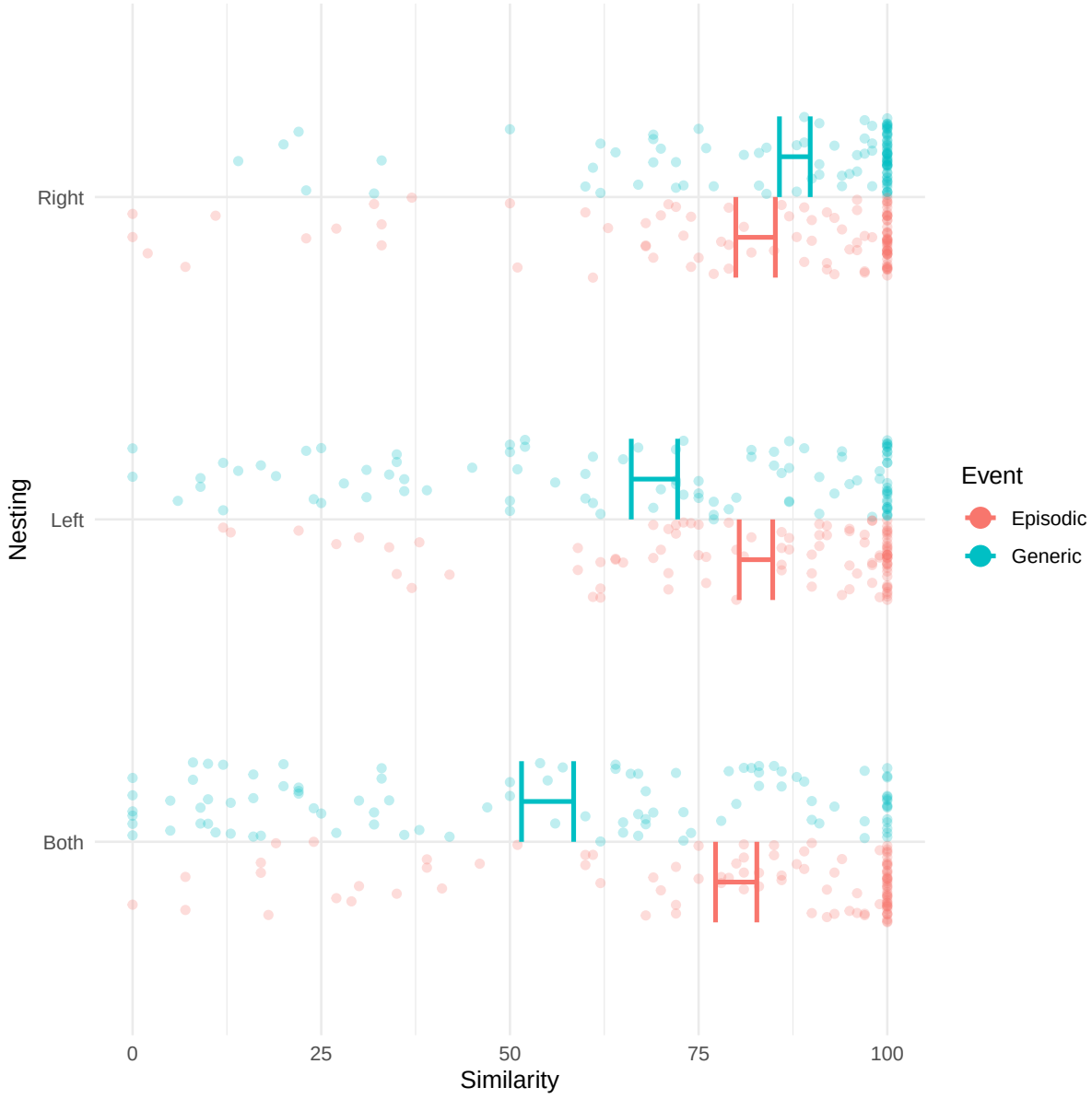


Figure 2

Similarity ratings collected in the experiment. Error bars represent ± 1 *standard error of the mean.*

(episodic vs. generic); Model 3, with main effects of both nesting and event; and model 4, with the addition of interaction between nesting and event.

Table 2
Summary of the models used in the data analysis to predict similarity ratings.

	Parameters	
Model name	Fixed effects	Random effects
Model 0 (Reduced)	None	Participant+Questionnaire
Model 1 (Nesting)	Nesting	Participant+Questionnaire
Model 2 (Event)	Event	Participant+Questionnaire
Model 3 (Main effects)	Nesting+Event	Participant+Questionnaire
Model 4 (Full)	Nesting*Event	Participant+Questionnaire

4.5. Results

The results of this analysis are shown in Table 3. The model comparison overwhelmingly favored Model 4, with an interaction between nesting and event, in both AIC and BIC (Model 4 $AIC = -486$, $BIC = -447$; Model 0 $AIC = -404$, $BIC = -387$; $BF_{40} = 1.2 \times 10^{13}$). In Bayes Factors, Model 4 was 1.7×10^4 times as likely as the next best candidate, Model 3, with no interaction term ($AIC = -458$, $BIC = -427$; $BF_{34} = 1.7 \times 10^4$).

Table 3
Comparisons between the models predicting the similarity between statements as reported by the participants.

Model name	AIC	BIC	BF_{X0}	BF_{4X}	
Model 0 (Null)	-404	-387	1	1.2×10^{13}	
Model 1 (Nesting)	-435	-409	5.8×10^4	2.1×10^8	
Model 2 (Event)	-420	-398	2.6×10^2	4.6×10^{10}	
Model 3 (Main effects)	-458	-427	7.0×10^8	1.7×10^4	
Model 4 (Full)	-486	-447	1.2×10^{13}	1	$\star$

$\star$ Best model:
BF_{X0}: Bayes Factor indicating how the given model (X, from 0 to 4) compares to the null model (0). The greater the number the better the model.
BF_{4X}: Bayes Factor indicating how much better the best model (4) was compared to the given model (X, from 0 to 4). For example, Model 4 is 17242 (1.7×10^4) times as likely as Model 3 to explain our data.

As noted above, the theory under evaluation predicts that left-nested and both-nested episodic conditionals should be judged more similar to their conjunctive counterparts than the matched generic conditionals. These predictions were confirmed. For left-nested conditionals, episodic conditionals were rated as more similar than generic conditionals, and the difference between conditions was highly significant ($p = .00054$ in a 100,000-sample

two-sided bootstrap t-test). For both-nested conditionals, the same pattern emerged, with ratings favoring the episodic variant and a highly significant difference ($p = 5.197 \times 10^{-8}$ in a 100,000-sample two-sided bootstrap t-test).

The theory also predicts that both right-nested generic and right-nested episodic conditionals should both be equivalent to their conjunctive counterparts. Consistent with this prediction, the difference in similarity ratings between these two conditional types was not significant ($p = .1188$ in a 100,000-sample two-sided bootstrap t-test).

4.5.1. Discussion

The experiment showed overwhelming evidence for an interaction between nesting type and generic vs. eventive interpretation. This result is as predicted by de Finetti when combined with linguistic theories of genericity. This theory predicts equivalence between nested episodic conditionals and their conjunctive counterparts across the board, but the same equivalence holds for generic conditionals only in the right-nested "GEN (If A, then if B, then C)".

In terms of the specific qualitative predictions outlined above, the results were as predicted by the de Finetti theory. When we look just at episodic conditionals, the nested conditionals were rated as highly similar to the conjunctive equivalents that the de Finetti theory predicts, across all three nesting types.

- Right-nested episodic: If A, then if B then C $\Leftrightarrow$ If A and B, then C
- Left-nested episodic: If B if A, then C $\Leftrightarrow$ If A and B, then C
- Doubly-nested episodic: If B if A, then D if C $\Leftrightarrow$ If A and B and C, then D

Right-nested generic conditionals were also rated as highly similar to their conjunctive counterparts, as predicted. In contrast, left-nested generic conditionals were significantly less similar to the matched conjunctive conditionals, compared to episodic conditionals of the same nesting type. Doubly-nested generic conditionals were judged as even less similar to their purported conjunctive equivalents. The greater magnitude of the difference in this condition is possibly due to the likely presence of two generic operators, rather than one, in the logical forms of the doubly-nested generic stimuli that we used (see section 4.2 above).

5. Conclusion

Generic interpretation can obscure the logic of conditionals. Our experimental results show that the presence or absence of a linguistically unmarked generic operator can have a dramatic effect on participants' judgments of equivalence between sentence types. Using linguistic manipulations involving specific time reference, we were able to control whether conditional stimuli were interpreted as generic or episodic, and so whether the results were affected by complex compositional interactions between the conditional antecedent and other elements of the sentence's interpretation. A similar experiment which included only generic stimuli might have concluded, with Douven (2016), that de Finetti's predictions about nested

conditionals are badly wrong: similarity between nested conditionals and their conjunctive counterparts was indeed quite low in two of three generic conditions. However, by attending to the linguistically well-motivated distinction between generic and episodic conditionals, our results provide surprising confirmation for de Finetti's predictions. Our results are consistent with the claim that the basic logic of conditionals is precisely as de Finetti predicted, and the apparent counter-examples were due to the generic confound.

In the extensive literature on reasoning with conditionals, there has been no systematic effort to control for the generic vs. episodic character of the stimuli. This introduces important confounds: the presence of generic operators that interact semantically with conditional antecedents can influence our intuitions about argument validity in ways that do not reflect the basic logic of conditionals. For example, we saw above that failure to attend to generic interpretation can lead to spurious counter-examples to basic principles like Modus Tollens (example (18) in section 3). This point opens up the worrying possibility that the use of generic stimuli may have led to incorrect conclusions about the logic of conditionals in previous studies. False negatives are an obvious worry: spurious counter-examples can lead to incorrect rejection of logical principles that are in fact valid. False positives are also possible: if a certain reasoning pattern is judged as valid in an experiment, but the stimuli have a generic interpretation, it is possible that the validity is due to an interaction between GEN and the conditional antecedent. The same pattern may not be judged valid when the experiment is repeated with episodic-only stimuli.

More broadly, this study highlights the importance of controlling for subtle linguistic factors when designing materials to test theories of reasoning. In this respect, the present study underlines the lessons of previous work that has brought key lessons from natural language semantics to bear on the study of reasoning with vague language (Kamp & Partee, 1995), quantifiers (Geurts, 2003), and modal expressions (Lassiter & Goodman, 2015). Decades of careful empirical and theoretical work on the lexical and compositional semantics of natural languages have uncovered a large body of findings that can enrich the study of reasoning and help researchers avoid drawing incorrect conclusions based on a superficial analysis of the linguistic materials involved.

Data availability

Experimental data and R code for the analyses can be found here: (removed for anonymous review)

References

Adams, E. W. (1975). *The logic of conditionals: an application of probability to deductive logic*. Springer.

Baratgin, J. (2021). Discovering early de finetti's writings on trivalent theory of conditionals. *Argumanta*, *6*(2), 267–291. https://doi.org/10.14275/2465-2334/202112.bar

Baratgin, J., Over, D. E., & Politzer, G. (2013). Uncertainty and the de finetti tables. *Thinking & Reasoning*, *19*(3-4), 308–328. https://doi.org/10.1080/13546783.2013.809018

Baratgin, J., Over, D. E., & Politzer, G. (2014). New psychological paradigm for conditionals and general de Finetti tables. *Mind & Language*, *29*(1), 73–84. https://doi.org/10.1111/mila.12042

Baratgin, J., Politzer, G., Over, D. E., & Takahashi, T. (2018). The psychology of uncertainty and three-valued truth tables. *Frontiers in Psychology*, *9*(1479). https://doi.org/10.3389/fpsyg.2018.01479

Belnap, N. D. (1970). Conditional assertion and restricted quantification. *Noûs*, *4*(1), 1–12. https://doi.org/10.2307/2214285

Berio, L., & Musholt, K. (2023). How language shapes our minds: on the relationship between generics, stereotypes and social norms. *Mind & Language*, *38*(4), 944–961. https://doi.org/https://doi.org/10.1111/mila.12449

Braine, M. D. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological review*, *85*(1), 1. https://doi.org/10.1037/0033-295X.85.1.1

Braine, M. D., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In *Psychology of learning and motivation* (pp. 313–371). Elsevier. https://doi.org/10.1016/S0079-7421(08)60365-5

Cantwell, J. (2008a). Changing the modal context. *Theoria*, *74*(4), 331–351. https://doi.org/10.1111/j.1755-2567.2008.00028.x

Cantwell, J. (2008b). Indicative conditionals: factual or epistemic? *Studia Logica*, *88*, 157–194.

Carlson, G. N. (2012). Habitual and generic aspect. In R. I. Binnick (Ed.), *The oxford handbook of tense and aspect*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195381979.013.0029

Carlson, G. N., & Pelletier, F. J. (Eds.). (1995). The generic book. Chicago: The University of Chicago Press.

Comrie, B. (1976). *Aspect: an introduction to the study of verbal aspect and related problems* (Vol. 2). Cambridge university press.

de Finetti, B. (1936). La logique de la probabilité. *Actes du congrès international de philosophie scientifique*, *4*, 1–9.

de Finetti, B. (1970). *Teoria delle probabilità: sintesi introduttiva con appendice critica*. Einaudi.

de Finetti, B. (1975). *Theory of probability: A critical introductory treatment* (A. Machí & A. Smith, Trans.). John Wiley & Sons.

de Finetti, B. (1979a). Ambiguïtés de jargon et de fond dans le domaine des probabilités. *Journal de la Société de Statistique de Paris*, *120*(2), 70–76.

de Finetti, B. (1979b). Jargon-derived and underlying ambiguity in the field of probability. *Scientia*, *114*, 713–716.

Douven, I. (2016). On de Finetti on iterated conditionals. *Computational models of rationality: Essays Dedicated to Gabriele Kern-Isberner on the occasion of her 60th birthday*, 265–279.

Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition*, *117*, 302–318. https://doi.org/10.1016/j.cognition.2010.08.015

Edgington, D. (1995). On conditionals. *Mind*, *104*(414), 235–329. https://doi.org/10.1017/CBO9780511753466

Egré, P., & Rossi, L. (Eds.). (2024). *Handbook of three-valued logics*. MIT Press.

Égré, P., Rossi, L., & Sprenger, J. (2021). De finettian logics of indicative conditionals part i: trivalent semantics and validity. *Journal of Philosophical Logic*, *50*, 187–213.

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

Farkas, D. F., & Sugioka, Y. (1983). Restrictive if/when clauses. *Linguistics and Philosophy*, 225–258.

Gelman, S. A. (2021). Generics in society. *Language in Society*, *50*(4), 517–532. https://doi.org/10.1017/S0047404521000282

Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, *86*(3), 223–251. https://doi.org/10.1016/S0010-0277(02)00180-4

Harman, G. (1972). Logical form. *Foundations of Language*, 38–65.

Huitink, J. (2008). *Modals, Conditionals and Compositionality* (Doctoral dissertation). Radboud University Nijmegen.

Joaquin, J. J. (2023). Weak Kleene and Other Weak Logics of Conditionals. *Notre Dame Journal of Formal Logic*, *64*(3), 281–290. https://doi.org/10.1215/00294527-2023-0008

Johnson-Laird, P. N. (2010). Against logical form. *Psychologica Belgica*, *50*(3), 193–221. https://doi.org/10.5334/pb-50-3-4-193

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, *57*(2), 129–191. https://doi.org/https://doi.org/10.1016/0010-0277(94)00659-9

Kratzer, A. (1991). Conditionals. In A. von Stechow & D. Wunderlich (Eds.), *Semantik: ein internationales Handbuch der zeitgenössischen Forschung* (pp. 651–656). Walter de Gruyter.

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: an introduction. In G. N. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 1–124). Chicago: The University of Chicago Press.

Lassiter, D. (2020). What we can learn from how trivalent conditionals avoid triviality. *Inquiry*, *63*(9-10), 1087–1114. https://doi.org/10.1080/0020174X.2019.1698457

Lassiter, D., & Baratgin, J. (2021). Nested conditionals and genericity in the de finetti semantics. *Thought: A Journal of Philosophy*, *10*(1), 42–52. https://doi.org/https://doi.org/10.1002/tht3.478

Lassiter, D., & Goodman, N. D. (2015). How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition*, *136*, 123–134. https://doi.org/10.1016/j.cognition.2014.10.016

Leslie, S.-J. (2017). The original sin of cognition. *The Journal of Philosophy*, *114*(8), 395–421.

Lewis, D. (1975). Adverbs of quantification. In E. L. Keenan (Ed.), *Formal semantics of natural language* (pp. 178–188). Cambridge University Press.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*(3), 297–315. https://doi.org/10.2307/2184045

Milne, P. (1997). Bruno de Finetti and the logic of conditional events. *The British Journal for the Philosophy of Science*, *48*(2), 195–232.

Mura, A. (2009). Probability and the logic of de Finetti's trievents. In M. C. Galavotti (Ed.), *Bruno de Finetti, radical probabilist* (pp. 201–242). College Publications.

Mura, A. (2011). Towards a new logic of indicative conditionals. *Logic and Philosophy of Science*, *9*, 17–31.

Mura, A. (2021). Bypassing Lewis' triviality results. a Kripke-style partial semantics for compounds of Adams' conditionals. *6*(2), 293–354. https://doi.org/10.14275/2465-2334/202112.mur

Over, D. E., & Baratgin, J. (2017). The "defective" truth table. *The thinking mind: A festschrift for Ken Manktelow*, 15–28.

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

Over, D. E., & Evans, J. S. B. T. (2003). The probability of conditionals: the psychological evidence. *Mind & Language*, *18*(4), 340–358. https://doi.org/10.1111/1468-0017.00231

Politzer, G., Jamet, F., & Baratgin, J. (2020). Children's comprehension of conditional requests. In S. Elqayam, I. Douven, J. S. B. T. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind a tribute to David E. Over* (pp. 161–177). Routledge. https://doi.org/10.4324/9781315111902-10

Politzer, G., Over, D. E., & Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning*, *16*(3), 172–197. https://doi.org/10.1080/13546783.2010.504581

Remijsen, B., Ismael, A., & Ayoker, O. G. (2024). Complex clauses and serialization in Shilluk [Chapter under review for *LD&C Special Publication No. 14: A Grammar of Shilluk*].

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, *11*, 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Stanley, J. (2000). Context and logical form. *Linguistics and philosophy*, *23*(4), 391–434. https://doi.org/0.1023/A:1005599312747

Stanley, J., & Szabó, Z. (2000). On quantifier domain restriction. *Mind & Language*, *15*(2-3), 219–261.

Vidal, M. (2014). The defective conditional in mathematics. *Journal of Applied Non-Classical Logics*, *24*(1-2), 169–179. https://doi.org/10.1080/11663081.2014.911540

von Fintel, K. (2004). A minimal theory of adverbial quantification. In H. Kamp & B. H. Partee (Eds.), *Context-dependence in the analysis of linguistic meaning* (pp. 137–175). Brill.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Penguin.

Yalcin, S. (2012). A counterexample to modus tollens. *Journal of Philosophical Logic*, *41*(6), 1001–1024. https://doi.org/10.1007/s10992-012-9228-4

Appendix: Experimental items

Generic Left 1

In a construction company, a new employee is getting a tour with a supervisor during their first day at work: The new employee:
"Why couldn't we use this material?"
The supervisor:
"If this material gets soft if it gets hot, then it is not suited for our purposes."
Compared to the supervisor's response, how similar is the meaning of the following sentence?
"If this material gets hot and soft, then it is not suited for our purposes."
Absolutely different Absolutely the same

Episodic Left 1

In a construction company, a new employee is getting a tour with a supervisor during their first day at work: The new employee:
"Why couldn't we use this material?"
The supervisor:
"If this material gets soft right now if it got hot during the morning, then our workers won't be able to use it this afternoon."
Compared to the supervisor's response, how similar is the meaning of the following sentence?
"If this material got hot during the morning and it gets soft right now, our worker won't be able to use it this afternoon."
Absolutely different Absolutely the same

Generic Left 2

A recruiter reports their interview with an airline pilot candidate to the admission board: The recruiter:
"What should we do with this candidate?"
The board:
"If a candidate gets angry if they don't sleep well, then they shouldn't be a pilot."
Compared to the board's response, how similar is the meaning of the following sentence?
"If a candidate doesn't sleep well and gets angry, then they shouldn't be a pilot."
Absolutely different Absolutely the same

Episodic Left 2

A recruiter reports their interview with an airline pilot candidate to the admission board: The recruiter:
"What should we do with this candidate?"
The board:
"If this candidate gets angry during today's interview if he didn't sleep well last night, then the committee should deny him the position at tomorrow's meeting"
Compared to the board's response, how similar is the meaning of the following sentence?
"If this candidate didn't sleep well last night and gets angry during today's interview, then the committee should deny him the position at tomorrow's meeting"
Absolutely different Absolutely the same

Generic Right 1

The organizer of a birthday party gives instructions about the course of the evening to their guests: A guest:
"I usually take my car to go back home"
The organizer:
"If you stay late at a party, then you should spend the night there if you get drunk."
Compared to the organizer's response, how similar is the meaning of the following sentence?
"If you stay late at a party and you get drunk, then you should spend the night there."
Absolutely different Absolutely the same

Episodic Right 1

The organizer of a birthday party gives instructions about the course of the evening to their guests: A guest:
"I usually take my car to go back home"
The organizer:
"If you stay late tonight, then you should spend the night here if you get drunk during the party."
Compared to the organizer's response, how similar is the meaning of the following sentence?
"If you stay late tonight and you get drunk during the party, then you should spend the night here."
Absolutely different Absolutely the same

Generic Right 2

A young lady asks her mayor how she will be able to vote: The young lady:
"How can I vote?"
The mayor:
"If you are 18 or older, then you can vote if you come to the polls."
Compared to the mayor's response, how similar is the meaning of the following sentence?
"If you come to the polls and you are 18 or over, then you can vote."
Absolutely different Absolutely the same

Episodic Right 2

A young lady asks her mayor how she will be able to vote: The young lady:
"How can I vote?"
The mayor:
"If you turned 18 this February, then you can vote if you come to the polls in March."
Compared to the mayor's response, how similar is the meaning of the following sentence?
"If you come to the polls in March and you turned 18 in February, then you can vote."
Absolutely different Absolutely the same

Generic Both 1

Two friends discuss their parents' reaction to their bad grades: Thomas:
"Do you think my mom is going to be mad?"
Victor:
"If your mom gets angry if you come home with a D, then she will get angry if you come home with an F."
Compared to Victor's response, how similar is the meaning of the following sentence?

"If you come home with a D and a F and your mom gets angry, then she will get angry."
Absolutely different Absolutely the same

Episodic Both 1

Two friends discuss their parents' reaction to their bad grades: Thomas:
"Do you think my mom is going to be mad?"
Victor:
"If your mom got angry last semester if you came home with a D back then, then she will get angry if you come home with an F today."
Compared to Victor's response, how similar is the meaning of the following sentence?
"If you came home with a D last semester and your mom got angry back then and you come home with an F today then she will get angry."
Absolutely different Absolutely the same

Generic Both 2

Two car designers are wondering about the fact that they did not improve the solidity of their cars: Marc:
"What would happen if this car hit another car?"
Oliver:
"If this car is dented if it hits a tree, then it will be dented if it hits another car."
Compared to Oliver's response, how similar is the meaning of the following sentence?
"If this car hits a tree and is dented and it hits another car, then it will be dented."
Absolutely different Absolutely the same

Episodic Both 2

Two car designers are wondering about the fact that they did not improve the solidity of their cars: Marc:
"What would happen if this car hit another car?"
Oliver:
"If this car was dented last year if it hit a tree at that time, then it will be dented tomorrow if it hits another car then."
Compared to Oliver's response, how similar is the meaning of the following sentence?
"If this car hit a tree last year and was dented at that time and the car hits another car tomorrow then it will be dented then."
Absolutely different Absolutely the same