

Conditionals, Context, and the Suppression Effect

Fabrizio Cariani
Department of Philosophy
Northwestern University

Lance J. Rips
Department of Psychology
Northwestern University

forthcoming in *Cognitive Science*

[initial submission: 1/16/2014; R&R: 4/29/2014; revised: 10/2/2014; accepted:
3/9/2015; final submission: 4/10/2015]

Keywords: Reasoning, conditional sentences, suppression effect, strict conditional, Stalnakerian pragmatics, modality, common ground.

Abstract

Modus ponens is the argument from premises of the form *If A then B* and *A* to the conclusion *B* (e.g., from *If it rained, Alicia got wet* and *It rained* to *Alicia got wet*). Nearly all participants agree that the modus ponens conclusion logically follows when the argument appears in this Basic form. However, adding a further premise (e.g., *If she forgot her umbrella, Alicia got wet*) can lower participants' rate of agreement—an effect called *suppression*. We propose a theory of suppression that draws on contemporary ideas about conditional sentences in linguistics and philosophy. Semantically, the theory assumes that people interpret an indicative conditional as a context-sensitive strict conditional: true if and only if its consequent is true in each of a contextually determined set of situations in which its antecedent is true. Pragmatically, the theory claims that context changes in response to new assertions, including new conditional premises. Thus, the conclusion of a modus ponens argument may no longer be accepted in the changed context. Psychologically, the theory describes people as capable of reasoning about broad classes of possible situations, ordered by typicality, without having to reason about individual possible worlds. The theory accounts for the main suppression phenomena, and it generates some novel predictions that new experiments confirm.

Conditionals, Context, and the Suppression Effect

1. Introduction

Hardly any form of argument is as intuitively compelling as modus ponens, the argument form: *If A then B; A; Therefore, B*. For example:

If an X is written on the board, a Y is written on the board. An X is written on the board, *Therefore*, a Y is written on the board.

The first premise of a modus ponens (MP) argument is a conditional (*if-then*) statement, composed of an *antecedent* (*an X is written on the board*) and a *consequent* (*a Y is written on the board*). The second premise is the same as the antecedent. The conclusion of the argument (the part after “therefore”), is the same as the consequent. In many experiments on deductive reasoning, participants decide whether the conclusion of arguments like this one must be true when the premises are true (or whether the conclusion logically follows from the premises). When participants have no prior beliefs about the truth of the individual sentences, nearly all of them judge that the conclusion of modus ponens follows. According to a review by Evans, Newstead, and Byrne (1993), between 89% and 100% of adult participants (across studies) endorse (the relevant instances of) modus ponens.

The intuitive correctness of modus ponens seems closely linked to the way we understand conditional sentences. Someone who claimed that *If there is an X on the board then there is a Y on the board* and *There is an X on the board*, but denied that *There is a Y on the board* would seem to lack a basic understanding of *if...then*. This relation between conditionals and modus ponens is so close that some have claimed modus ponens to be constitutive of the meaning of conditionals (e.g., Boghossian, 2000; Dummett, 1991, pp. 272-274).¹

This strong connection, as well as the nearly unanimous endorsement of modus ponens in reasoning experiments, makes it surprising that some simple changes can cause people to withhold assent to arguments that resemble the Basic modus ponens form. Compare the following pair of arguments:

Argument 1: If she has an essay to write, she will study late in the library. She has an essay to write. *Therefore*, She will study late in the library.

Argument 2: If she has an essay to write, she will study late in the library. If the library stays open, she will study late in the library. She has an essay to write. *Therefore*, She will study late in the library.

Byrne (1989) first showed that, although virtually all participants thought that the conclusion of Argument 1 follows from its premises, only 38% thought the conclusion of Argument 2 does. This decrease in agreement is known as the *suppression effect*—in this case, suppression of modus ponens. The second argument differs from the first just by adding the second conditional statement about the library staying open. Since Argument 2 contains Argument 1, in the sense of containing its premises and conclusion, Byrne’s finding raises the question of how the new conditional manages to suppress the intuitive correctness of the modus ponens inference.

Experiments that demonstrate suppression have sometimes asked participants to decide, not whether the conclusion logically follows from the premises, but simply how confident they are in the conclusion. This wording may encourage participants to use inductive reasoning based on their background beliefs about the topic (in the case of the library example, beliefs about how libraries operate). Current research suggests that people perform inductive and deductive reasoning using distinct mental processes (Heit & Rotello, 2010; Markovits & Handley, 2005; Parsons & Osherson, 2001; Rips, 2001; Rotello & Heit, 2009), and if so, it is tempting to write off the suppression effect as not engaging deductive reasoning. We show later, however, that suppression is robust: It occurs even when participants are under standard deduction instructions to determine whether the conclusion logically follows from the premises.

Our positive theoretical aim in this article is to frame the phenomenon of suppression of modus ponens in the context of contemporary research in philosophy of language and semantics, and to defend a cognitive explanation of the phenomenon in terms of that account. Our approach does not rely on a nonstandard notion of deductive validity or on a probabilistic relation between premises and conclusion, as do many other proposals, but rather on independently motivated principles concerning the dynamics of conversation.

Crucially, however, the interaction between semantics and psychology goes in both directions: just as current research in semantics provides new ways of thinking about conditionals, the psychological results provide new case studies and new desiderata for a theory of meaning of the conditional. These new constraints will motivate us to fill out the semantic picture in new ways.

2. The suppression data

Say that an argument is *suppressed* when the addition of new premises decreases its rate of endorsement. In our earlier example, adding the second conditional premise (*If the library stays open, she will study late in the library*) suppresses the original modus ponens. In the following, we use *vindicated* as an antonym of sorts

for suppressed. We say that an inference is vindicated if the addition of new premises increases or leaves unchanged its endorsement rate.

In classical logic, adding premises to a deductively valid argument cannot make the argument invalid. Systems with this property are called *monotonic*. Systems in which adding premises can flip an argument from valid to invalid are *nonmonotonic*. For this reason, the phenomenon of suppression, applied to valid arguments, has sometimes been taken to show that some aspects of human deductive reasoning are better modeled by a nonmonotonic than a monotonic system (Stenning & van Lambalgen, 2005, 2007). In this section, we begin by looking at previous evidence for the suppression of modus ponens and modus tollens. These arguments provide the main evidence for nonmonotonicity. We then consider suppression of two invalid argument forms: affirming the consequent and denying the antecedent.

2.1. Modus Ponens data

As mentioned, the core of data about the suppression effect involves variations on modus ponens. We illustrate some of these variations with examples drawn from an experiment that we describe later in this section, but the examples are representative of stimuli that appear throughout the literature on the suppression effect. (Many of them come from earlier experiments.) In its simplest or Basic form, modus ponens consists of just a major conditional premise, a minor premise, and a conclusion. For example:

- (1) **Basic MP:** If it was raining, Alicia got wet. It was raining. *Therefore:* Alicia got wet.

We call the argument form containing only the major premise, the minor premise, and the conclusion, *Basic modus ponens*. As noted, a large proportion of participants endorse it. In studies specifically devoted to the suppression effect, the median percentage of endorsements for Basic modus ponens is also quite high, as shown in Fig. 1. In the figure, modus ponens arguments appear as solid light-colored bars, with Basic modus ponens at the far left.

The proportions cited here and in Fig. 1 come from suppression experiments that ask participants to decide explicitly whether a given conclusion follows (or ask them to produce a conclusion that follows) from the premises. (Thus, we exclude studies that ask for confidence ratings rather than judgments of validity.) We also limit consideration to arguments with two or three premises. The experiments we include in these tallies are Byrne (1989, Experiment 1); Byrne, Espino, and Santamaria (1999, Experiments 1-4); Chan and Chua (1994, using the percentages reported in Byrne et al., 1999); Dieussaert, Schaeken, Schroyens, and d'Ydewalle (2000, Experiment 1); Politzer (2005, Experiment 2); and Stevenson and Over (1995, Experiment 1). We refer to these studies as *explicit suppression experiments* in what follows. (We compare these with *implicit* suppression experiments in the General Discussion.)

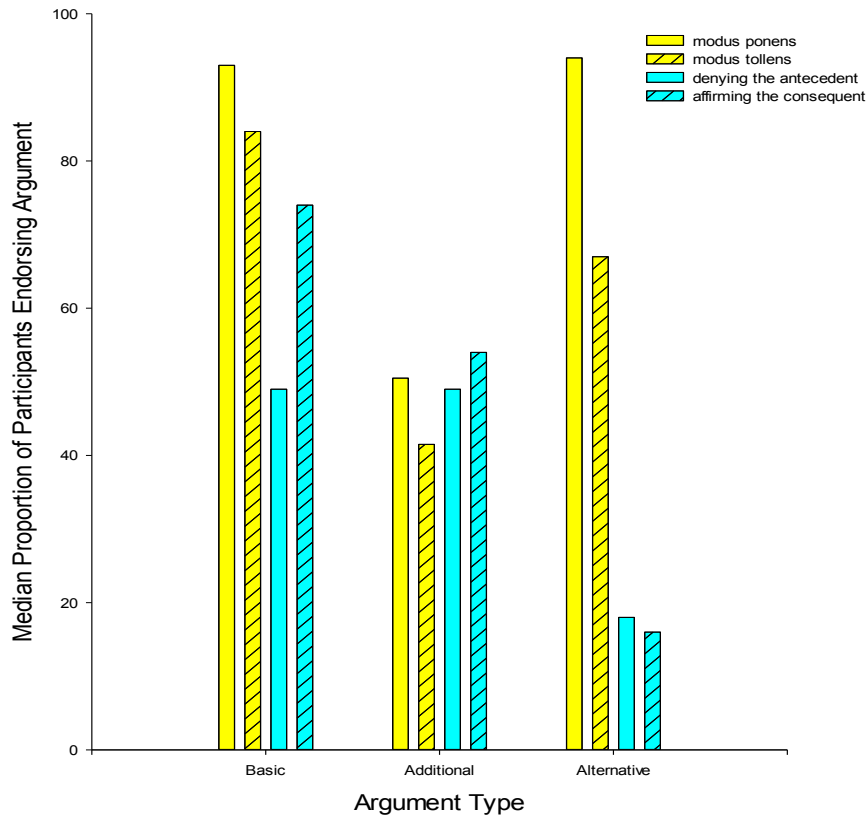


Figure 1. Percentage of endorsement of inference types over previous studies of the suppression effect (Byrne, 1989, Experiment 1; Byrne, Espino, & Santamaria, 1999, Experiments 1-4; Chan & Chua, 1994, using the percentages reported in Byrne et al., 1999; Dieussaert, Schaeken, Schroyens, & d'Ydewalle, 2000, Experiment 1; Politzer, 2005, Experiment 2; and Stevenson & Over, 1995, Experiment 1).

Other variations on modus ponens, however, produce noticeably lower endorsement rates. This is true, for example, of the following argument:

- (2) **Additional MP:** If it was raining, Alicia got wet. If Alicia forgot her umbrella, Alicia got wet. It was raining. *Therefore*, Alicia got wet

The middle premise in (2) (“If Alicia forgot her umbrella, Alicia got wet”) is often referred to as an *Additional* premise, and we call arguments containing this premise *Additional modus ponens*. As Fig. 1 shows, only about half of participants endorse this type of argument. A similar suppression effect can also be achieved with a different logical structure for the second premise. We use here what we call a *Contravening* conditional (aka *disabling* conditional), such as:

- (3) **Contravening MP:** If it was raining, Alicia got wet. If Alicia remembered her umbrella, Alicia did not get wet. It was raining. *Therefore*, Alicia got wet.

The middle premise of (3) raises a possibility in its consequent (that Alicia did not get wet) that contravenes the possibility of the first premise's consequent (that she did get wet). The observation here (De Neys, Schaeken, & d'Ydewalle, 2003) is that it does not matter whether the second conditional premise directly raises the contravening possibility (e.g., as in (3)) or whether it raises its negation (as in (2)).

The conditional premises of Contravening modus ponens are similar to what philosophers in the literature on counterfactual conditionals call *Sobel sequences* (Lewis, 1973). These are pairs of counterfactual conditionals of the form *If A then C* and *If A and B then not C*. For example, *If it had rained, Alicia would have gotten wet* and *If it had rained and Alicia had remembered her umbrella, Alicia would not have gotten wet*. This similarity will play a role in our discussion.

Additional and Contravening modus ponens both produce suppression. They contrast with yet another type of inference—one in which an extra conditional does *not* interfere with modus ponens. This conditional is often referred to as an *Alternative* premise. An example appears as the middle premise of:

- (4) **Alternative MP:** If it was raining, Alicia got wet. If she fell into the pond, Alicia got wet. It was raining. *Therefore*, Alicia got wet.

Alternative modus ponens patterns with the Basic form and vindicates modus ponens. Across the explicit suppression experiments, cited earlier in this section, the median rate of endorsement for Alternative modus ponens is nearly the same as for Basic modus ponens, as shown in Fig. 1. Byrne (1989) has suggested that the disanalogy between modus ponens with Alternative and Additional premises reveals that people's inclination to perform modus ponens inferences depends on the content and not merely on the logical form of the arguments. That is, the logical form of both Alternative and Additional modus ponens would seem to be *If A then B; If C then B; A; Therefore, B*. Yet Alternative modus ponens receives many more endorsements than Additional modus ponens.

2.2. Further data points

The literature on suppression is not limited to modus ponens. *Modus tollens* (MT) is the pattern of argument from *If A then B* and *Not B* to *Not A*. It is well known in the psychology of reasoning that a significant portion of participants do not endorse Basic instances of modus tollens, such as:

- (5) **Basic MT:** If it was raining, Alicia got wet. Alicia did not get wet. *Therefore*, It was not raining.

Endorsement rates range from 41% to 81% across studies in which the arguments are not associated with prior beliefs, according to the review by Evans et al. (1993).

The percentage is slightly higher in experiments devoted to suppression (the explicit suppression experiments, cited in Section 2.1), as shown in Fig. 1 (light-colored hashed bars).

Importantly, this number *decreases* when Additional premises are present, but doesn't significantly decrease with Alternative premises. The pattern is therefore similar to what we found with modus ponens: Additional premises suppress both modus ponens and modus tollens, whereas Alternative premises vindicate both. This naturally suggests that suppression applies to modus tollens much as it does to modus ponens (compare the two light-colored bars in the figure).

A few experiments on suppression have also studied related argument types known as *affirming the consequent* and *denying the antecedent* (e.g., Byrne, 1989; Byrne et al., 1999; Dieussaert et al., 2000; Rumin, Connell, & Braine, 1983). Basic examples of affirming the consequent have the form *If A then B; B; Therefore, A*, and basic examples of denying the antecedent have the form *If A then B; Not A; Therefore, Not B*. These types of argument are traditionally known as "formal fallacies": Unlike modus ponens and modus tollens, they are generally not taken to be deductively valid. For present purposes, however, what is important about affirming the consequent and denying the antecedent is that Additional and Alternative conditionals have the opposite effects on these arguments than they have on modus ponens and modus tollens. For example, the following arguments illustrate affirming the consequent with an Additional premise in (6) and an Alternative premise in (7):

(6) **AC Additional:** If it was raining, Alicia got wet. If Alicia forgot her umbrella, Alicia got wet. Alicia got wet. *Therefore*, It was raining.

(7) **AC Alternative:** If it was raining, Alicia got wet. If she fell into a pond, Alicia got wet. Alicia got wet. *Therefore*, It was raining.

The data from suppression experiments on affirming the consequent and denying the antecedent are much thinner than for the other inference types we have discussed, and we should therefore be cautious in interpreting the results. However, Alternative forms, such as (7), tend to suppress both these arguments, whereas Additional forms, such as (6) tend to vindicate them. Fig. 1 shows this trend, where the darker-colored solid bars represents denying the antecedent and the darker-colored hashed bars represents affirming the consequent. Clearly, the Additional forms receive much higher endorsement rates than the Alternative forms, a pattern that reverses that for modus ponens and tollens.

We note that investigators have obtained effects similar to those in Fig. 1 in experiments in which participants see only the Basic forms of the conditional arguments. In these studies, the conditionals vary in the extent to which they bring to mind contravening or alternative possibilities (e.g., Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys et al., 2003; Geiger & Oberauer, 2007; George, 1995; Thompson, 1994). For example, participants may suppress the Basic modus

ponens argument in (1) or the Basic modus tollens argument in (5) if they think of situations in which it is raining but someone does not get wet. Investigators have interpreted these results as suggesting that participants take into account prior knowledge about the conditional in deciding whether to endorse the argument—for example, whether the antecedent is sufficient or necessary for the consequent in everyday settings. As we discuss later (see the General Discussion), however, these experiments produce more variable outcomes than those that manipulate arguments by adding explicit (Alternative or Additional) premises, and they pose special interpretive issues. In this article, we focus on arguments in which the Alternative or Additional premises appear explicitly and in which participants receive standard instructions to evaluate an argument for deductive validity (e.g., to determine whether the conclusion must be true when the premises are true).

2.3. Two new studies of suppression

A review of previously published studies on suppression, such as those summarized in Fig. 1, revealed several open issues that motivate our proposal. We therefore conducted two new experiments that we designed to resolve these questions. We describe the methodology of these experiments here, reporting the results when they become relevant in later sections.

Most of the psychological literature has dealt with conditionals of the form *If event, WILL consequence*. For example, *If it is raining, Alicia will get wet* has the form of these *do/will* conditionals (the label is from Bennett, 2003). Some philosophers have argued that *do/will* conditionals are *subjunctive* (or counterfactual) conditionals in disguise (Bennett, 2003, pp. 13-16). The statement just mentioned would seem to be true just in case it would also be true at some later time to say, *If it had been raining, Alicia would have gotten wet*. This last sentence is explicitly subjunctive. What this shows about *do/will* conditionals isn't entirely clear. However, since we are interested in testing suppression as a phenomenon concerning *indicative* conditionals rather than subjunctives, we carried out an initial experiment of our own using two groups of participants: One group received *do/will* conditionals, and a second group received comparable conditionals in the past tense, conditionals that are *bona-fide* indicatives. The conditional premises in our sample arguments (e.g., the first premise of (1) and the first two premises of (2), (3), and (4)) are all past-tense indicatives.

Our first experiment used four different versions of modus ponens, including a Basic form ((1) is an example of a stimulus item) and a Contravening form (e.g., (3)). We used Contravening modus ponens instead of Additional modus ponens in order to compare related conditions, as we discuss later; earlier studies have found both these argument types to suppress modus ponens (see Section 2.1). We also included two further kinds of arguments that we label *Conjunctive modus ponens* and *Disjunctive modus ponens* (note that these labels applied to different argument forms in Rips, 1994).

Conjunctive MP: If it was raining, then Alicia got wet. If it was raining and Alicia remembered her umbrella, Alicia did not get wet. It was raining. *Therefore,* Alicia got wet.

Disjunctive MP: If it was raining, then Alicia got wet. If Alicia remembered her umbrella, Alicia did not get wet. It was raining. *Therefore,* Either Alicia remembered her umbrella or Alicia got wet.

The Conjunctive form was similar to the Contravening argument but included the antecedent of the first premise as a conjunct in the antecedent of the second premise. The Disjunctive form was also similar to the Contravening argument, but the conclusion was a disjunction of the antecedent of the second premise and the consequent of the first. We included Conjunctive and Disjunctive modus ponens here because they help motivate the theoretical approach we describe in the third section of this article.

We phrased each of the four argument types in 12 different content variations, such as the one about Alicia getting wet. Most of these content versions came from earlier experiments on suppression. Participants read all 48 arguments, and we presented these, one-at-a-time, on a computer screen in a new random order for each participant. For each argument, the participants were under instructions to “ask yourself whether the last of these statements (the one below the line) follows logically from the others. By ‘logically follows’ we mean that if the statements above the line are true, the statement below the line must be true. Click on either ‘logically follows’ or ‘does not logically follow’ to record your decision.”

We also conducted a second experiment that followed the same procedure as the first. The differences concerned the arguments that appeared as stimuli. In this second study, we included exactly the same Basic and Conjunctive modus ponens arguments as in the first. However, we altered the Contravening form by replacing the Contravening conditional with a premise that asserted that its antecedent “might” occur. For example, the Contravening argument in (3) was replaced by: *If it was raining, Alicia got wet; Alicia might have remembered her umbrella; it was raining; [therefore,] Alicia got wet.* Similarly, we replaced the second conditional in the Disjunctive forms with a “might” claim. In place of the Disjunctive MP argument above, for example, participants saw: *If it was raining, then Alicia got wet; Alicia might have remembered her umbrella; it was raining; [therefore,] either Alicia remembered her umbrella or Alicia got wet.* In this second study, all conditionals were past-tense indicatives.

Participants were introductory psychology students at Northwestern University who took part in the studies to fulfill a course requirement. We excluded three participants from further analysis since they reported studying logic in one of their college classes. Of the remaining participants, 15 were in the do/will condition and 16 in the past tense condition of Experiment 1, and 16 were in Experiment 2.

Although the general shape of the suppression effect is apparent in Fig. 1, an explicit theory of suppression is more elusive. Byrne (1989) spearheaded an explanation in terms of mental models. Pragmatic accounts of the phenomenon followed soon thereafter. More recent work has focused on developing probabilistic and nonmonotonic theories. In the next three sections, we develop our own account. After doing so, we compare it with extant theories, noting the principal elements of agreement and disagreement.

3. The meaning of conditionals

We aim to explain the suppression data by appealing to a combination of familiar theses about the meaning of conditionals and the dynamics of linguistic context. Our approach to ideas concerning presuppositions and pragmatics largely builds on Stalnaker's work (1970, 1973, 1976, 1984). Our account, however, does not incorporate his semantics for the conditional. Moreover, we add principles that are not part of the standard Stalnakerian framework. Our motivation for this approach is the belief that the development of formal semantics in the last 30 years has uncovered patterns of explanation that remain untapped in the psychology of reasoning.

With this in mind, Sections 3 and 4 describe our basic tools; Section 5 articulates the explanation of suppression. We start by reviewing some formal semantic theories of the conditional. We zero-in on an analysis of the conditional that has acquired major prominence among philosophers and linguists, but whose psychological implications have not been explored. We then ask whether these views in formal semantics have implications for the psychology of reasoning. Specifically, how should we interpret the apparatus of possible worlds that is so fundamental to the semantics? We adopt the view that people construct analogues of the possible-worlds model by thinking in terms of suppositions and suppositional reasoning. We then move to some assumptions about the pragmatics of conversation. Although psychological experiments are not, strictly speaking, conversations, we believe that established models of linguistic pragmatics can be fruitfully brought to bear on questions concerning how information is stored and processed by actual reasoners, as many investigators have previously argued (e.g., Hilton, 1995; Schwarz, 1996; Sperber, Cara, & Girotto, 1995).

An initial example will help preview our account. The philosophical literature on conditionals has noted the failure of a variant on Conjunctive MP:

- (8) (i) If it was raining, then Alicia got wet. (ii) If it was raining and Alicia remembered her umbrella, Alicia did not get wet. (iii) It was raining and Alicia remembered her umbrella. *Therefore*, Alicia got wet.

The premises appear to be collectively consistent (i.e., not contradictory), but if modus ponens is valid, they seem to entail a contradiction. By (ii) and (iii), Alicia did

not get wet. But (iii) entails that it was raining, so by (i) she did get wet. Lycan (1993, 2001) declared modus ponens invalid on the basis of similar examples.

An alternative, and in our view preferable, explanation is that, as we go through the inference, we are tacitly shifting the background of assumptions relative to which we evaluate the conditionals.² When we accept (i), we implicitly accept that Alicia did not remember her umbrella; but, to entertain (ii), we must abandon this assumption. Somewhere along the way, we switched contexts, and modus ponens is only guaranteed to work in a fixed context. Our hypothesis is that a detailed implementation of this idea suffices to account for the experimental findings from the suppression effect.

3.1. Four analyses of the conditional

The story of modern theories of the conditional begins with the material conditional analysis. For purposes of comparison, we can write its definition as follows (from now on we will use *(if A)(C)* for the indicative conditional to stay neutral about its nature):

Material: *(if A)(B)* is true iff *A* is false or *B* is true

It is useful to think about what it takes for a conditional to be *false* on a given analysis. In this case, *(if A)(B)* is false if and only if *A* is true and *B* is false. Although the material conditional has a glorious history (and some prominent supporters, e.g., Jackson, 1991, and Williamson, 2007), research in the last half century has largely been the history of its decline. For example, as Stalnaker (1975) points out, under the material conditional the negation of *(if A)(B)* is *A & not B*. Yet, it seems incorrect to say that someone who denies *If the gardener didn't do it, the butler did* must accept *The gardener didn't do it* (see also Edgington, 1996, for a battery of similarly inspired counterexamples).

One of the early competitors to Material was the Strict Conditional (Lewis, 1918), which we'll call here Canonical Strict to distinguish it from a more modern variant that will be of more direct interest to us. To state the definition of the canonical strict conditional, we need the notion of a *possible world*. We set aside foundational questions about the nature of possible worlds, and regard each world, for purposes of sketching semantic theories, as some abstract object that determines a truth-value for all the atomic sentences (e.g., *It was raining*; *Alicia gets wet*; *Alicia forgot her umbrella*) in the formal language we use to represent our statements. In one such world, for example, it rained and Alicia got wet; in another, it rained and Alicia did not get wet; and so on. Given this notion, the canonical strict conditional has the following truth condition:

Canonical Strict: $(if A)(B)$ is true at world w iff for every A -world v , B is true at v ,

where by A -world we mean a world in which A is true. The key innovation is not the relativization to worlds *per se*, but the fact that the truth of $(if A)(B)$ at a world w can depend on what goes on at worlds that are different from w . Again, it helps to think about how a conditional can be false on this analysis: $(if A)(B)$ is false at w just in case some world v (possibly distinct from w or possibly identical to it) makes A true but B false.

Canonical Strict resolves some of the material conditional's problems, but one may reasonably worry that it is a bit *too* strict. Consider:

(9) If the match is struck, it will light.

In many everyday settings, (9) is true. But possibilities exist, perhaps rather remote ones, in which the match is struck, but it does not light. Maybe aliens have tampered with the match or stray cosmic rays have invisibly altered its phosphorus content. Remote possible worlds that include these bizarre situations render the canonical strict conditional false in our own world. There are three main approaches to this problem: (i) probabilistic accounts, (ii) *variably strict* accounts, and (iii) contextualist variants of the strict analysis.

On a probabilistic account, (9) is acceptable because the associated conditional probability is high, even though it may be less than 1. As we mentioned, probabilistic accounts are important players in the literature on the suppression effect, but our aim here is to spell out a qualitative model of the phenomenon, and, for that, we rely on a qualitative model of the conditional. Further comments on the probabilistic approach appear later in this section and in the General Discussion.

The variably strict approach (see Stalnaker, 1968; Lewis, 1973, although Lewis did not adopt it for indicative conditionals) evaluates conditionals relative to a possible world w and a similarity relation s between worlds. Here is a simplified version of the semantics for the special case in which there is a maximally close world (which Lewis denies).

Variably Strict: $(if A)(B)$ is true at w,s iff for every world v that is maximally close to w (according to s) and such that A is true at v , B is true at v .

Variably Strict solves the problem posed by (9) by evaluating the context in the *most similar* world (or worlds) in which the match is struck. It does not matter if there is a distant possibility in which the match is struck but does not light. One problem with Variably Strict conditionals, however, is that they fail to imply the equivalence of $(if A \& D)(B)$ with $(if A)((if D)(B))$. For instance, *If it rained and Alicia did not remember her umbrella, she got wet* would seem to mean the same as *If it rained, then if Alicia did not remember her umbrella, she got wet*.

For this and other reasons, one of the more strongly supported analyses in recent work on conditionals is a context-sensitive variant of the strict conditional analysis. According to this variant, the conditional behaves like a strict conditional

but within a contextually restricted range of possible worlds. The view is brought up, and quickly dismissed, as a theory of counterfactuals in Lewis (1973) and entertained by Stalnaker (1984), who suggests that there might be little to adjudicate between it and the Variably Strict account:

One can defend a strict conditional account of conditionals against the counterexamples and arguments we have given by emphasizing the context-dependence of conditionals. One may argue that the conditional is *semantically* a fixed strict conditional but that the domain of possible worlds relative to which it is defined varies with context. (Stalnaker, 1984, p. 125)

This approach has been gaining support in recent literature on counterfactual conditionals (von Fintel, 2001a; Gillies, 2007). Moreover, in the case of bare indicative conditionals, a strict conditional approach is essentially equivalent to the seminal analysis of conditionals as restrictors given by Kratzer (1991).

To translate it to a formal semantics, we evaluate conditionals relative to a possible world w and a state i . Loosely speaking, the state consists of the possible worlds that are relevant in a given context (we must say more about the nature of i , and we will do so shortly). The state i is generally interpreted as a set of worlds compatible with a salient set of beliefs or items of knowledge.

(Contextualized) Strict: $(if A)(B)$ is true at w, i iff for every A -world v in i , B is true at v, i .

We will refer to this simply as *Strict*. In a vivid sense, *Strict* is intermediate in strength between *Material* and *Canonical Strict*. For *Material*, a conditional is true if and only if it is not the case that A is true and B is false in the actual world. For *Strict*, a conditional is true if and only if it is not the case that A is true and B is false in any one of a contextually determined range of possibilities.³ For *Canonical Strict*, we drop the restriction to contextually determined possibilities: Any possibility can undermine a conditional, no matter how farfetched.

What we call “*Strict*” is a simplified account that is sufficient for our presentational needs. A more complex variant assigns a double role to the conditional antecedent (Gillies, 2009; Yalcin, 2007, 2012). Let $i+A$ be the result of eliminating from i all those worlds in which A is false.

Strict*: $(if A)(B)$ is true at w, i iff for every world v in $i+A$, B is true at $v, i+A$.

This definition is invoked to deal with the interaction between modals (such as *would* or *might*) and conditionals (Gillies, 2009). This sort of difference between *Strict* and *Strict** does not matter to our examples, since none of them contains an overt modal in its consequent. What is important about these formulations for our purposes is that they directly incorporate context dependence through the i

parameter, and context dependence is, we will claim, a key aspect of the suppression phenomena. If we do use Strict*, however, we can make some important conceptual connections. First, define:

i accepts *A* iff for all worlds *v* in *i*, *A* is true at *v*,*i*.

With this notion of acceptance (Yalcin, 2007, 2012), we can also rewrite our clause for Strict*.

Strict* (equivalent definition): (*if A*)(*B*) is true at *w*,*i* iff *i*+*A* accepts *B*.

This way of rewriting Strict* reveals some close parallels with probabilistic accounts. In a probabilistic theory, the conditional probability *Pr* (*consequent* | *antecedent*) determines the acceptability of the conditional. This is the probability of the consequent when we rule out situations in which the antecedent is not true. Likewise, Strict* ties the acceptability to the fact that the conditional consequent has a certain property after we rule out the worlds that do not satisfy the antecedent. For probabilistic accounts, this property is quantitative and contributed by the salient probability model. For Strict*, this property is qualitative, and contributed by the salient state *i*.

Strict* resolves the problem raised by (9) by assuming that the worlds in which the match is struck but does not light are outside the relevant state *i*. In a typical context in which someone utters (9), *i* will not include possible worlds in which something outlandish interferes with the match's lighting.⁴

3.2. From formal semantics to the psychology of reasoning

A central goal of formal semantics is to characterize the conditions in which sentences of a language are judged acceptable (or unacceptable) by competent speakers, based on the sentences' meaning. There is a clear gap between this task and the task of formulating a model of how people reason with the information those sentences express. In particular, strict conditional theories invoke possible worlds to explain the meaning of conditionals, as we have just seen. But how can human reasoners grasp possible worlds?

This gap might be bridged in several ways. First, we might take the semantic theories as supplying blueprints for the design of psychological analogues. Although the semantics deploys possible worlds, we might provide a psychological model in which talk of possible worlds is reinterpreted. For example, Rips and Marcus (1977) produce a psychological analogue of Variably Strict in a framework that replaces talk of possible worlds with talk of suppositions—finite sets of mental sentences that represent how a hypothetical situation differs from the actual one. Suppositional theories of conditionals are also discussed in Braine and O'Brien (1991) and Evans et al. (2005). In a different representational style, Johnson-Laird and Byrne (1991, 2002) use mental models to depict sets of possibilities.

A second view might be to say that possible-worlds analyses can be used to provide formal *characterizations* of certain classes of inferences that people are disposed to make. The representation does not involve the claim that participants reason through the semantic analysis. Clearly, participants will not be able to carry out all the inferences that come out as valid according to the abstract representation, but it might be possible for the representation to provide a computational-level description (in the sense of Marr, 1982) of the nature of human inferences and of their appropriateness. Along these lines, Partee (1989, p. 117) suggests that “...possible-world semantics can just as readily be viewed as a theory of the content of psychological states associated with the asserting or accepting of various sentences, though not directly as a theory of the mechanisms employed by the language users themselves.”

Here, we remain neutral on this second option, but we take the semantic analysis to suggest a psychological analog (though one that differs from previous suppositional theories, such as those cited earlier). We call this analog a *scoreboard* (adopting a metaphor from Lewis, 1979), and it integrates both semantic and pragmatic aspects of an argument. As we will see, the scoreboard allows people to reason about a context without having to represent individually each of the possible worlds it contains.⁵

4. The dynamics of conversation

Like any semantic analysis, Strict does not stand alone. It is complemented by a pragmatic theory that details how utterances of various kinds can affect the context of a conversation or of an episode of reasoning. We now turn to a description of our commitments in this arena.

4.1. General pragmatics

Our approach employs a discourse model along the lines of Stalnaker’s (1970, 1973, 1976, 2002). At the core of Stalnaker’s model is the notion of *presupposition*. For him, presupposing is, in the first instance, something that participants to a conversation *do*—namely, *taking some proposition for granted*. The *common ground* is the set of propositions that are taken to be true for the purposes of the conversation (and mutually recognized to be so). Clark (1996) has developed a closely related notion of common ground as a central part of a theory of human collaboration, especially the collaboration people exhibit in conversations. On Clark’s account, common ground is the information people share in conversational settings, and changes to the common ground—for example, additions or corrections of this information—are the result of joint actions. We therefore view common ground as an independently needed aspect of communication and take advantage of this concept in understanding the way people interpret conditionals in context.

The propositions in the common ground determine a set of possible worlds—the ones that are compatible with the common ground. Stalnaker calls the set of these possibilities *the context set*. Each possibility in this set is such that all propositions in the common ground are true. Fig. 2 illustrates this terminology. The

figure depicts the possible worlds $w_1, w_2, \dots, w_k, w_{k+1}, \dots$ in the context set c , each containing all the propositions in common ground (the p_i 's), as well as additional propositions that distinguish one possible world from another. Although common ground constrains the context set, it usually leaves open propositions that lie beyond the conversation's current topic. We will see, however, that these further propositions can play an important role in reasoning.

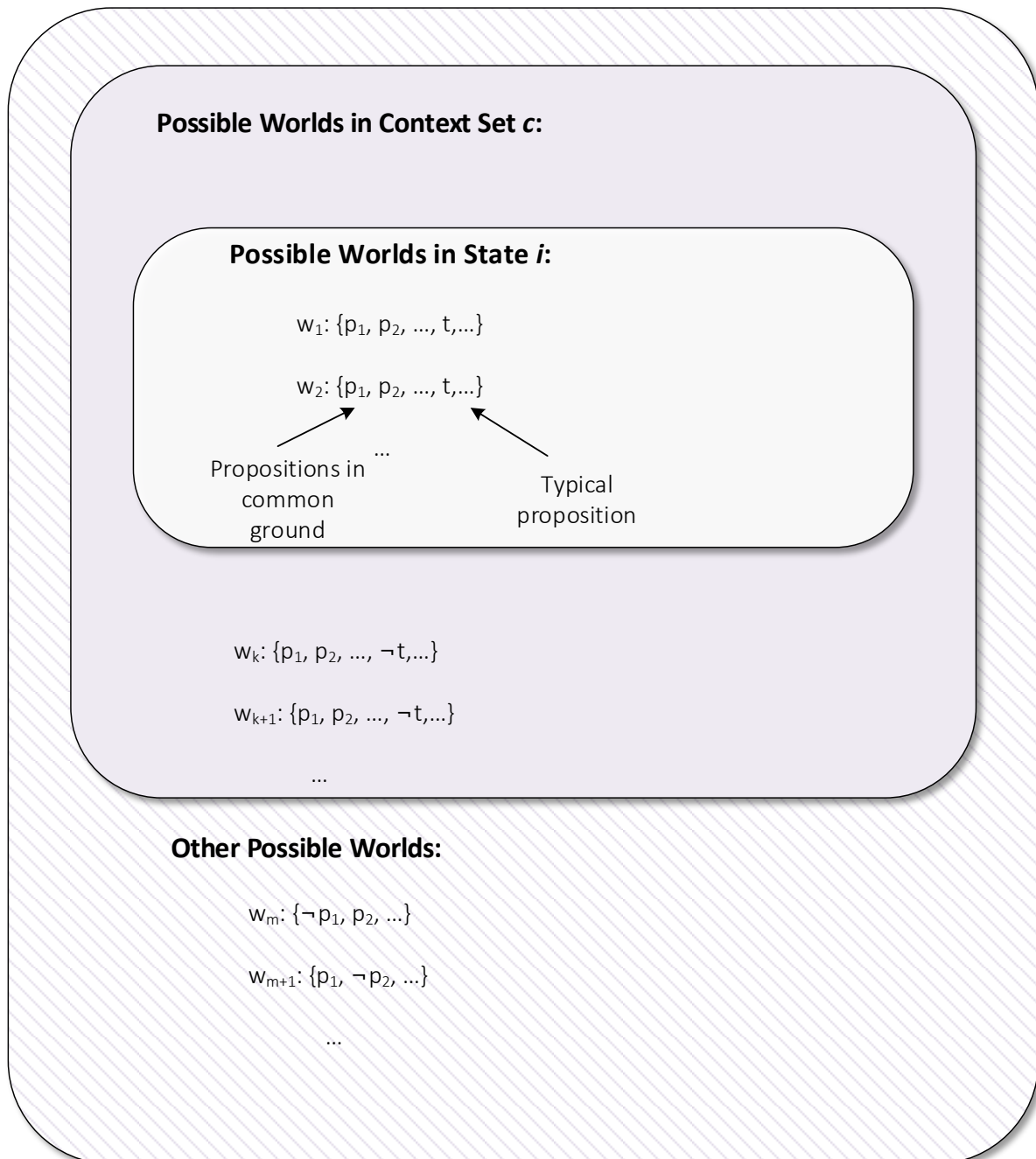


Figure 2. The space of possible worlds relevant to the interpretation of an indicative conditional. Worlds in the context set c are those consistent with propositions in the common ground (p_1, p_2, \dots). Worlds in state i are those directly invoked in interpreting the conditional.

The context set allows us to define a notion of acceptance: A proposition p is accepted in context C just in case it is true at every possible world in the context set of C . For example, Fig. 2 shows proposition p_i as accepted. A successful assertive utterance of a sentence A in context C adds the proposition expressed by A (in C) to the common ground. Putting together these two ideas:

Acceptance: Successful assertion of a sentence A requires the resulting context set to exclude those possibilities at which A is false.

For reasons we will come to shortly, assertion of conditionals follows a somewhat different principle; so we apply Acceptance only to nonconditional assertions (we might have to entertain further deviations, for instance in the case of modals).

On our account, the common ground will play a further role. It helps us constrain the state i that $(if A)(B)$ quantifies over according to Strict. Suppose that we are in a context C with c as its context set. Then we assume:

Inclusion: The state i —needed for evaluating an indicative conditional $(if A)(B)$ —must be a subset of the context set c .

Notice that we do not *identify* states with context sets. They only have to be subsets of the latter, as shown in Fig. 2. Inclusion guarantees that when a successful assertion updates the common ground, it will also affect the states of further conditional statements in the discourse.

In the semantics literature, determination of parameters like i is frequently left up to context. Although this might satisfy some of the goals of semantic theory, it is not enough for our present purposes: To get genuine predictions about reasoning, we need a more predictive account of the state i (Knobe & Szabo, 2013, make a similar point in a related theoretical context). We will assume that reasoners make tacit assumptions that the state initially includes typical properties and excludes atypical properties and properties ruled out by background knowledge. The typical proposition t in Fig. 2 expresses these properties. This idea accords with well-known evidence that reasoners often rely on representative possibilities rather than exhaustively considering all possibilities (e.g., Kahneman & Tversky, 1972). The next section describes the role this psychological assumption plays in producing the suppression effect.

Although the state i is initially set in this typical way, it can evolve on the basis of independent principles. For example, we stipulate the following presupposition:

Possibility: To evaluate an indicative conditional $(if A)(B)$ in context C , the state i must be compatible with A .

In this setting, *evaluating a conditional* means computing the way in which it affects the context. Since we follow a broadly Stalnakerian outlook, we suppose that violation

of presuppositions like Possibility results in a defective context. By *compatible*, we mean that the state must contain at least one possible world in which the proposition that *A* expresses is true. If the contextually supplied state *i* for the conditional does not satisfy Possibility, we hypothesize that *i* expands a bit to accommodate it. Together with Inclusion, Possibility entails:

Context-Set Possibility: Evaluating an indicative conditional of the form $(if\ A)(B)$ presupposes that *A* be possible within the context set.

Stalnaker (1975) proposed thinking of indicative $(if\ A)(B)$ as presupposing that *A* is compatible with the common ground. For him, the mark of counterfactual conditionals is that they are not bound by this presupposition. We are concerned here with indicative conditionals, so we will adopt the view that $(if\ A)(B)$ presupposes that *A* is possible within the context set.

It is important, however, that the state cannot easily expand beyond the common ground. This is required to explain why the following discourse sounds odd (Willer, 2013b).

- (10) Joe was in Austin. If Joe was not in Austin, he was in Dallas.

(10) is dangerously close to a contradiction. This feeling can be explained if we think that the first sentence updates the common ground with the information that Joe was in Austin, while the second requires of the common ground that it be compatible with the possibility that Joe was not in Austin.

4.2. Pragmatics of suppression

The pragmatic ideas we have discussed up to this point are fairly standard. We now add a few further commitments that are essential to our model of the suppression effect.

First, we assume that possibility assumptions can interact: Suppose we must add to the state some possibilities in which *A* is true in accord with the Possibility principle in order to evaluate the conditional $(if\ A)(B)$. Suppose also that there is a salient division within the context between *B*-worlds and $\neg B$ -worlds. In such a case, unless there are contextually important reasons against this, the state *i* expands to make possible both *A* & *B* and *A* & $\neg B$. We call this constraint *Composite Possibility*. This constraint does not stem from the presuppositions of any specific sentence. It is motivated, rather, by the idea that once certain possibilities are brought to salience, their combinations with other contextually relevant possibilities are also made salient. Suppose, for example, that the current context contains worlds in which it is raining and others in which it isn't. If we then consider the conditional *If Alicia remembered her umbrella, then she didn't get wet*, we should include both remember-and-rain worlds and remember-and-not-rain worlds in the new state.⁶

Our second non-standard commitment is to a principle also related to Possibility.

Negative Possibility: To evaluate an indicative conditional (*if A*)(*B*) in context *C*, the state *i* must be compatible with $\neg A$.

The motivation behind Negative Possibility is that if *A* were settled in *i*, a speaker who was in a position to assert (*if A*)(*B*), would also be in a position to assert the stronger *It must be the case that B*. For Gricean reasons, then, asserting (*if A*)(*B*) would signal that the speaker envisions some possibilities in which $\neg A$.⁷ For example, in asserting *If it was raining, Alicia got wet*, we invoke a state containing worlds in which it was *not* raining, as well as worlds in which it was.

The last piece of our account concerns the effects on the common ground of utterances of bare unembedded conditionals (i.e., conditionals with non-modal consequents that are also not parts of larger sentences). One might think that such utterances must update the common ground by ruling out the worlds in which the conditional is false. The problem is that, as we defined it, **Strict** implies that (*if A*)(*B*) is either true at every world (if *i* lacks *A* & $\neg B$ worlds) or it is false at every world (if *i* contains *A* & $\neg B$ worlds). It would follow that when we rule out from the context set the worlds at which (*if A*)(*B*) is false, we either rule out every possible world or none. This does not seem to be a plausible story about the effect of bare conditional utterances. Instead, we want successful assertive utterances of (*if A*)(*B*) to have the following effect:

Exclusion: Accepting an utterance of (*if A*)(*B*) has the effect of ruling out those worlds in the state *i* in which *A* is true but *B* is false.

So accepting *If it was raining, Alicia got wet*, excludes from *i* raining-and-not-wet worlds. The Exclusion principle takes up the slack left by our exempting conditionals from the Acceptance principle. An important technical problem in the philosophy of language is how to give a pragmatic story that unifies these two principles, but we will not tackle this problem here.⁸ We limit ourselves to taking conditionals to have a *sui generis* effect on the common ground—an effect that is captured by Exclusion.

5. Explanations and predictions

We are ready to argue that our assumptions about conditionals and contexts predict that suppression of modus ponens and tollens is a plausible response to the introduction of Additional and Contravening premises. We then extend the account to deal with suppression of the “fallacies” (affirming the consequent and denying the antecedent).

The general template of our explanation is as follows: Assume that we start with some context C that determines a context set c and a state i . Evaluating premises generally affects i in the systematic ways we have specified. Suppose that i^* is the state after all premises have been interpreted. We predict that participants will accept the conclusion as following just in case it is true at every world in i^* . Fig. 3 summarizes the routine that participants use in deciding whether to endorse a conclusion, and it puts together the principles that we have discussed in the previous sections. According to this picture, people consider the premises of an argument in sequence, using the content of the premises to expand the current state (by Composite Possibility or by use of the modal *might*) or contract the current state (by Acceptance or Exclusion). The horizontal tracks of the model represent the different processes required for interpreting conditional premises (*if A*)(*B*), premises of the form *Might*(*P*), and minor premises (e.g., *A* or \neg *B*). Once all the premises have been interpreted, reasoners decide whether the conclusion follows, according to the criterion we just described (shown on the right of Fig. 3). In the next few subsections, we illustrate how this procedure accounts for the suppression phenomena.

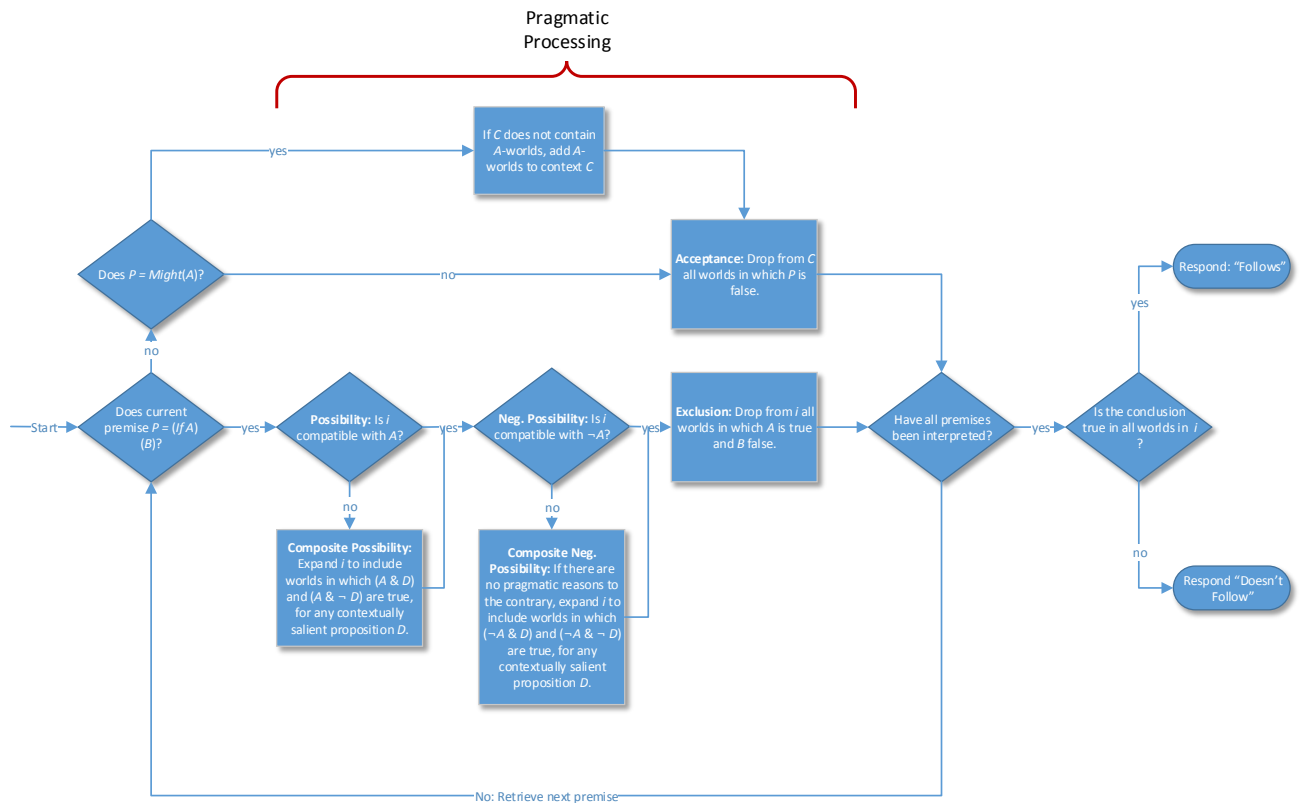


Figure 3. Steps in determining whether the conclusion of a conditional argument follows from its premises. Pragmatic steps appear under the central bracket; semantic steps elsewhere. See text for details on the individual steps in this process. Note that both A and B must be non-modal sentences.

5.1. The scoreboard representation and basic MP/MT

We will use a diagram, which we call a scoreboard, to represent the constraints imposed on context. For example, take the case of Basic modus ponens:

- (11) If it was raining, Alicia got wet. It was raining. *Therefore*, Alicia got wet.

We start by first representing how this argument structures the context set. The conditional premise of (11) calls to salience four types of possibilities:

Basic MP		
	wet	¬wet
rain		
¬rain		

The cells in the scoreboard represent different kinds of possibilities without representing individual possible worlds. For example, the shaded cell at the upper-left represents all possible worlds in the context set c in which it was raining and Alicia got wet. In this diagram and the following ones, a cell is shaded if and only if the state i contains at least one such world. (Recall that state i is the subset of possible worlds in context set c that is relevant to evaluating the conditional; see Fig. 2.) At the beginning of the problem, on first inspecting the conditional, all four types of possibilities are compatible with the initial state, assuming that people have no preconceived ideas about the truth of these sentences.

In introducing the scoreboard, we are not committing ourselves to the idea that people mentally represent states and context sets in a literal diagram. We could implement the scoreboard as a set of mental sentences or as a diagrammatic representation or in some other way. Our predictions do not depend on the details of the format. What is important is the idea that we can reason about states and context sets in a summary fashion that specifies the possibilities they contain without having to enumerate these possibilities one-by-one. The scoreboard simply provides a convenient and neutral means to track changes to the state in response to new premises. Of course, further research may call for further specification of the scoreboard, but the current suppression data do not seem to require it.

As people reason through the argument, possibilities are eliminated (by accepting various premises) or added (by processing pragmatic constraints such as Possibility). For example, in Basic modus ponens, processing the major premise of (11) eliminates from the state those possibilities in which rain&¬wet (by Exclusion). The minor premise of (11) then rules out the ¬rain possibilities. Given a name S for a sentence in an argument, we write x_s to mean that interpreting S rules out a certain type of world from the state. For example, in Basic modus ponens we have:

Basic MP		
	wet	\neg wet
rain		\times _{major}
\neg rain	\times _{minor}	\times _{minor}

This scoreboard illustrates why Basic modus ponens is accepted. Once all the premises are interpreted, the only possibilities that are compatible with the state are in the rain & wet cell. Reasoners should accept the conclusion of the argument, *Alicia got wet*, as following from the premises since the final state rules out the possibility (\neg wet) that she stayed dry.

As Fig. 1 indicates, nearly all participants in traditional experiments on suppression endorse Basic modus ponens. Fig. 4 (left-most bars) shows that the same is true in our own experiments. On 99% of trials in Experiment 1 and 92% in Experiment 2, participants agreed that the modus ponens conclusion logically followed from its premises. As Fig. 4 also suggests, this was the case both when the arguments contained past-tense indicative conditionals, such as the major premise of (11), as well as when they contained the more traditional do/will conditionals (e.g., *If it rains, Alicia will get wet*). In fact, the overall results in Experiment 1 showed no difference due to the wording of the conditional ($F(1,29) = 0.13$, $MS_e = 0.066$, $\eta_p^2 = .004$, $p > .10$) and no interaction between wording and argument type ($F(3,87) = 0.16$, $MS_e = 0.039$, $\eta_p^2 = .005$, $p > .10$). We will therefore collapse over the wording variable in reporting the results in the rest of this article.

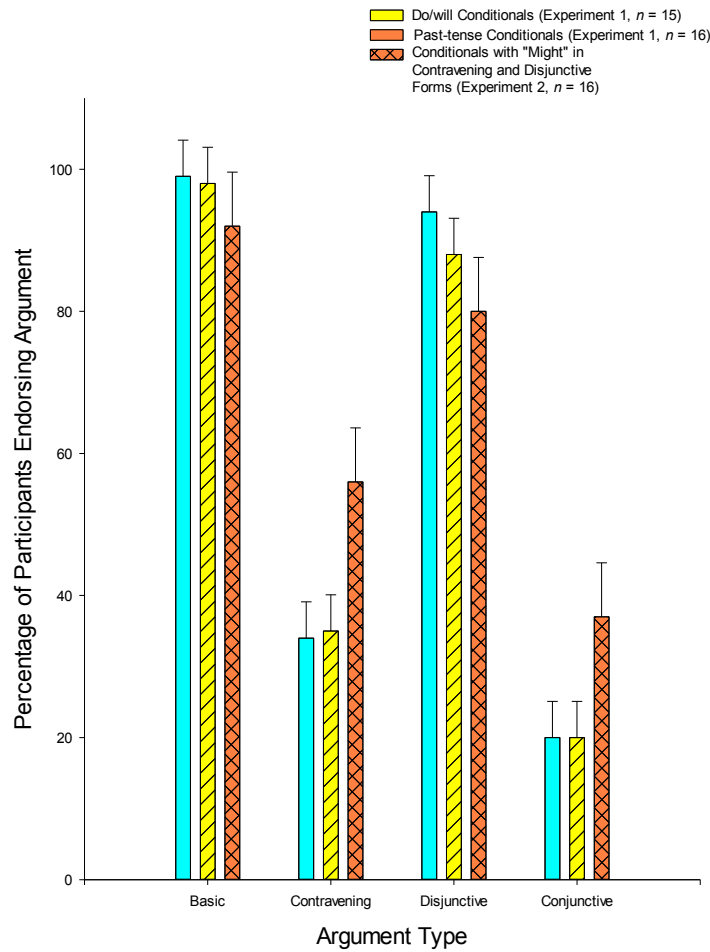


Figure 4. Percentage of endorsements for four argument forms, Experiments 1 and 2. See text for a description of the arguments' structure.

Moving on to modus tollens, we predict the Basic form to be accepted for similar reasons. Consider, for example, *If it was raining, Alicia got wet; Alicia did not get wet; therefore, it was not raining.*

Basic MT		
	wet	¬ wet
rain	\times_{minor}	\times_{major}
¬rain	\times_{minor}	

After interpreting the major premise, the state may contain rain possibilities. But they are all rain & wet possibilities. Once the minor premise (e.g., *Alicia did not get wet*) rules out all the wet possibilities, no rain possibilities are left.

As we mentioned earlier, many experiments on conditional reasoning have found that participants endorse basic modus ponens more frequently than basic modus tollens. In suppression experiments, however, this difference is quite modest (< 10 percentage points in the studies summarized in Fig. 1). We leave open here just how our proposal might be supplemented to yield this result, assuming that it is a reliable one. One option is to suppose that ruling out possibilities incompatible with a negated statement is a more complex mental operation than ruling out possibilities compatible with a positive statement. Another is to suppose that the constraints on context that we represent via diagrams are actually to be given a more complex representation and that reasoner's dispositions to endorse a given inference depend on the complexity of the deductions that will take the reasoner from a set of constraints on context to a given conclusion—the idea being that the deduction associated with modus tollens is a more complicated task than the deduction associated with modus ponens. See Braine, Reiser, and Rumin (1998) and Rips (1994) for proposals along those lines.⁹

5.2. *Contravening MP*

We begin our account of suppression with Contravening modus ponens.

- (12) If it was raining, Alicia got wet. If Alicia remembered her umbrella, Alicia did not get wet. It was raining. *Therefore*, Alicia got wet.

To analyze Contravening modus ponens, we need to follow the evolution of context more slowly than we did for Basic modus ponens. To start, we can assume that the first premise sets up the same initial scoreboard that we displayed for the Basic modus ponens case, and in evaluating this premise, we eliminate those possibilities according to which rain & ¬wet, just as before.

However, mention of remembering an umbrella in the antecedent of the second premise brings to mind some additional possibilities. The context set is initially compatible with seven such possibilities, and the scoreboard expands to accommodate them (according to the Composite Possibility principle). In this representation, we need to distinguish the state *i* (again, consisting of the shaded cells) from the context set. The unmarked white cells represent worlds that belong to the context set, but not to *i*.

Contravening MP (step 1)	remember		¬remember	
	wet	¬ wet	wet	¬ wet
rain				\times_{major}
¬ rain				

Notice that this last scoreboard is asymmetric. In enlarging the scoreboard, we've retained the analysis of the major premise on the right, as consistent with what happens if Alicia does not remember her umbrella. But the scoreboard is so far uncommitted to what happens if she remembers it. In this way, the diagram incorporates the assumption that not remembering the umbrella is more typical than remembering it in a situation in which the first premise is true. We suppose that the sentences involved in the argument, as well as the preceding context, might help fix the dimensions of typicality. In the example, we suppose that because Alicia's remembering her umbrella is brought up as a possibility by the Contravening premise's antecedent, remember/¬remember is a salient dimension of typicality.

At this point, we need to evaluate the full Contravening premise *If Alicia remembered her umbrella, Alicia did not get wet*. The problem is that the sentence violates Possibility (see Section 4.1). By Possibility, the state (shaded region in the scoreboard) ought to include some remember-worlds. Which ones? Several different types of worlds might be thrown in at this step, but in accord with our typicality assumption, we add those remember-worlds that are not ruled out by background knowledge. Following this criterion, we introduce three new kinds of possibilities:

Contravening MP (step 2)	remember		¬remember	
	wet	¬ wet	wet	¬ wet
rain				\times_{major}
¬ rain				

We will assume that people do not include possibilities, such as ¬rain & wet & remember worlds, that are contrary to ordinary experience, but this does not affect our predictions concerning Contravening modus ponens. What is essential is that the expansion in Step 2 introduces possibilities in which rain & remember & ¬wet.

Finally, we finish processing the Contravening and Minor premises. The minor premise rules out all the worlds in which it does not rain (by the Acceptance principle), while the contravening conditional appropriately rules out worlds in which rain & remember & wet (by Exclusion).

Contravening MP (step 3)	remember		¬remember	
	wet	¬ wet	wet	¬ wet
rain	χ_{contra}			χ_{major}
¬ rain		χ_{minor}	χ_{minor}	χ_{minor}

The upshot is that, in the final scoreboard, two types of worlds are in the state: rain & ¬wet & remember and rain & wet & ¬remember. The conclusion of the argument is again *Alicia got wet*, but the possibility remains that she did *not* get wet (if she remembered her umbrella).

Experiment 1 confirms that participants find Contravening modus ponens unconvincing. Overall, endorsement rates differed significantly among the four types of arguments in this experiment (Basic, Contravening, Disjunctive, and Conjunctive modus ponens), $F(3, 87) = 124.04$, $MS_e = 0.039$, $\eta_p^2 = .810$, $p < .001$. In particular, Fig. 4 shows that participants endorsed Contravening modus ponens on only 35% of trials, significantly less often than they endorsed Basic modus ponens (99%), according to a Tukey HSD test ($\alpha = .05$).

This completes our explanation of why Contravening modus ponens is suppressed. However, our diagram above provides three additional insights. First, the scoreboard predicts that Disjunctive modus ponens should *not* be suppressed. Disjunctive modus ponens has the same premises as Contravening modus ponens, but with the conclusion *Either Alicia remembered her umbrella or Alicia got wet*. Even though the final state does not support the conclusion that Alicia got wet, it does support the conclusion that either she remembered her umbrella or she got wet. Indeed, this is precisely what we see in Fig. 4. Participants endorsed the Disjunctive form on 91% of trials. This percentage is not significantly different from that of Basic modus ponens, but does differ from that of Contravening modus ponens by the Tukey HSD test.

A second bonus of the analysis concerns Conjunctive modus ponens. Recall that this is the argument whose middle premise is: *If it was raining and Alicia remembered her umbrella, Alicia did not get wet* instead of *If Alicia remembered her umbrella, Alicia did not get wet*. The antecedent of this new premise makes explicit the rain & remember possibilities that were only implicit in the Contravening argument. We would therefore expect a suppression effect at least as great as that for Contravening modus ponens. To our knowledge, no previous tests of this prediction appear in the literature, but Fig. 4 shows that the results from Experiment 1 confirm it. Participants accepted Conjunctive modus ponens on only 20% of trials in that experiment. This percentage is significantly less than that for even Contravening modus ponens by the same analysis mentioned earlier. In sum, the Tukey HSD test shows that modus ponens and Disjunctive modus ponens do not differ significantly. However, both have

significantly higher endorsement rates than Contravening modus ponens, which in turn has significantly higher rates than Conjunctive modus ponens.

The third advantage of the analysis concerns why participants report reinterpreting the conditionals as “If was raining *and* Alicia forgot her umbrella...” (Stenning & van Lambalgen, 2007). The idea is that the context we reach after we evaluate all the premises of the Contravening argument includes possibilities in which Alicia remembered to take her umbrella. Hence, expressing how the major premise affected the context requires clarifying that in processing the major premise we only ruled out worlds in which Alicia forgot her umbrella.

One last remark: Even though in Contravening modus ponens, suppression is the majority judgment, it is not a universal judgment. About a third of participants do endorse modus ponens in Contravening inferences. This raises the question: *Why do some participants not suppress modus ponens?* We conjecture that these participants either use different criteria for typicality (in particular, they do not assume that not remembering is more typical than remembering) or do not include possibilities in which rain & \neg wet & remember when they expand the state (in Step 2 above). Either of these operations allows modus ponens arguments to go through.

5.3. Additional MP and MT

Additional modus ponens replaces the contravening premise of (12) with:

- (13) If Alicia did not remember her umbrella, she got wet.

The first steps in the evaluation of Additional modus ponens are identical to those of Contravening modus ponens.

Additional MP (step 1)	remember		\neg remember	
	wet	\neg wet	wet	\neg wet
rain				\times major
\neg rain				

As before, the state expands when we process the second conditional premise, (13); however, this time the state does not expand by Possibility, but rather by Negative Possibility. The antecedent of (13) has been incorporated in the state at the beginning, since not remembering the umbrella is the typical case in which the first premise is true. But Gricean reasoning suggests that a speaker of (13) also has in mind possibilities in which Alicia might have remembered her umbrella. The state therefore expands to include some worlds in which the antecedent is not true.

Additional MP (step 2)	remember		¬remember	
	wet	¬ wet	wet	¬ wet
rain				X _{major}
¬ rain				

Next, the Additional premise is interpreted by ruling out the worlds in the state that are compatible with the antecedent and the negation of the consequent, in accord with the Exclusion principle.

Additional MP (step 3)	remember		¬remember	
	wet	¬ wet	wet	¬ wet
rain				X _{major} , X _{additional}
¬ rain				X _{additional}

Finally, in the last step, we process the minor premise.

Additional MP (step 4)	remember		¬remember	
	wet	¬ wet	wet	¬ wet
rain				X _{major} , X _{additional}
¬ rain		X _{minor}	X _{minor}	X _{additional} , X _{minor}

As before, we predict that modus ponens is suppressed because possibilities in which rain & ¬wet & remember appear in the final state. Again, we hypothesize that participants who do accept Additional modus ponens either employ different criteria for typicality or do not expand the context at Step 2 to include rain & ¬wet & remember worlds.

Notice, too, that we could eliminate the critical rain & ¬wet & remember possibilities by giving participants arguments in which the minor premise specified not only that it rained but also that Alicia did not remember her umbrella (i.e., *It was raining* would be replaced by *It was raining and Alicia did not remember her umbrella*).

This extra information should vindicate modus ponens, since the only remaining possibilities are now rain & wet & \neg remember worlds. Byrne (1989, Experiment 2) used just such arguments and found 100% acceptance of these items, in line with this prediction. In the same vein, our analysis also provides an explanation of what would seem at first to be a puzzling response to Additional arguments. When participants are asked to say what follows from the premises of such arguments, they sometimes respond with the equivalent of *If she did not remember her umbrella, she got wet* (Byrne et al., 1999, Table 5). But this conclusion repeats the argument's Additional premise (see (13)), and it is odd to think that participants would find it informative merely to restate a premise as a conclusion. However, a glance at the last of our scoreboards shows that, at the end of the argument, this response might be considered informative after all, since it calls attention to some of the remaining possibilities (rain & wet & \neg remember worlds) in which Alicia did get wet.

The explanation of Additional modus tollens is not significantly different from that of Additional modus ponens, and diverges only at Step 4. In Additional modus tollens, the minor premise (*Alicia did not get wet*) rules out wet -possibilities. But as we argued, the state includes a possibility in which rain & \neg wet & remember, and this is not ruled out by the last premise. This possibility means that participants should balk at drawing the conclusion (*It was not raining*).

At this point, we have worked through enough details to consider an objection to our account. We seem to incorrectly predict that reversing the order of the premises in Basic MP should generate suppression. Suppose one is given premises A and $(if A)(B)$ in that order. The first premise should rule out the $\neg A$ -worlds. By Negative Possibility, the second premise should first reinstate some $\neg A$ -worlds, and then eliminate the $(A \& \neg B)$ worlds, leaving open some $(\neg A \& \neg B)$ -worlds. Hence, we should not be able to conclude B . However, Reversed MP (i.e. A , $(if A)(B)$, therefore B) is about as robust as Basic MP (Giroto, Mazzocco, & Tasso, 1997).

To meet this objection, we must tell a more nuanced version of our story. We oversimplified in presenting Negative Possibility as a mechanical effect on the common ground, on a par with, say, Acceptance. However, Negative Possibility is justified by Gricean reasoning, and we should not generally expect the effect of Gricean reasoning to be so univocal. Uttering a conditional $(if A)(B)$ in a context that settles the truth of A violates a Gricean norm. But there are at least two possible responses to this violation, one more appropriate to Additional MP and the other to Reversed MP.

In Additional MP, it seems plausible to interpret the violation as signaling that some tacit presuppositions must be revised. This does not look plausible in Reversed MP: after all, the $\neg A$ -worlds are ruled out in the very same stretch of discourse by a premise that has just been uttered. Following the Gricean reasoning in detail suggests a natural alternative interpretation that is appropriate to this second case: Recall that the motivation of Negative Possibility is that if $\neg A$ -worlds are ruled out, the speaker should, other things equal, prefer an utterance of *Must(B)* to an utterance of $(if A)(B)$. However, one might yet utter the conditional to signal that acceptance of B is

dependent on acceptance of *A* rather than an independently established feature of the context. Of course, for all we have said, this is also an available upshot of the Gricean violation in some cases of Additional MP. But notice that the logic of our argument does not require that Negative Possibility be the *only* possible upshot of the Gricean violation. It is enough for our purposes if, upon encountering the violation, some reasoners interpret it by revising some tacit presuppositions and others interpret it by revising along the lines we just suggested. In fact, this would also explain why the suppression effect is stronger for Contravening MP (which turns on Possibility, which is ordinarily viewed as a presupposition) than it is for Additional MP.

5.4. Epistemic MP

Presuppositions and implicatures are not the only way to introduce the possibility that Alicia might remember her umbrella and not get wet. We believe that simply drawing attention to the possibility that Alicia might remember her umbrella can add rain & remember & \neg wet to the state. We predict, then, that we should encounter suppression if we replace the Additional or Contravening conditional with a non-conditional sentence that simply expresses the salient possibility claim:

- (14) Alicia might have remembered her umbrella.

Although we have not discussed the effect of *might* on the state, an intuitively plausible view is that (14) expands the state *i* to include remember possibilities (Willer, 2013a): in particular, in our case, to include rain & \neg wet & remember. This expansion leads to suppression in ways similar to the one we sketched above.

We tested this prediction in Experiment 2, and the results confirm it (see, also, Politzer, 2005, who used an extra premise similar to *It is not certain that Alicia forgot her umbrella*). The percentage of endorsements for the arguments in Experiment 2 appears as the double-hashed bars in Fig. 4. The first two points correspond to the Basic modus ponens argument and the argument in which the normal Contravening conditionals were replaced with *might* assertions, such as (14). As predicted, the results produced clear suppression: from 92% agreement with Basic modus ponens to 56% agreement for the arguments with *might*. The overall difference among the arguments was significant in Experiment 2 ($F(3,45) = 10.57$, $MS_e = 0.092$, $\eta_p^2 = .413$, $p < .001$), as was the difference between the Basic form and the Contravening form with *might* (by a Tukey HSD test, $\alpha = .05$). The HSD test showed that Basic modus ponens had higher endorsement than the Contravening form with *might* and the Conjunctive form. Similarly, the Disjunctive argument with *might* had higher endorsement than the Conjunctive form. There were no other significant differences. Although the argument with *might* suppresses modus ponens, it appears to produce less suppression than the normal Contravening form in Experiment 1 (see Fig. 4). However, the results from Experiment 2 are, in general, less extreme than those of Experiment 1, including the results for the Basic and Conjunctive forms, which were unchanged between the experiments. Hence, not too much should be read into these differences.

5.5. *Alternative MP and MT*

Alternative antecedents are importantly different from both Additional and Contravening ones. Recall our example of an Alternative premise:

- (15) If she fell into the pond, Alicia got wet.

Our account provides a reason why Alternative conditionals lead to the vindication of modus ponens and tollens.

The crucial point is that on the relevant dimension of typicality, we can assume that Alicia's not falling into the pond is more typical than Alicia's falling into the pond, which means that the initial scoreboard look like this:

Alternative MP, MT (step 1)	pond		¬pond	
	wet	¬wet	wet	¬ wet
rain				x_{major}
¬ rain				

After similar updates to the one we described above, the final scoreboard for modus ponens is:

Alternative MP (final)	pond		¬pond	
	wet	¬ wet	wet	¬ wet
rain		$x_{\text{alt.}}$		x_{major}
¬ rain	x_{minor}	$x_{\text{alt.}}$ x_{minor}	x_{minor}	x_{minor}

So the conclusion that Alicia will get wet is accepted. The final scoreboard for Alternative modus tollens also turns out to vindicate this argument:

Alternative MT (final)	pond		¬pond	
	wet	¬ wet	wet	¬ wet
rain	x_{minor}	$x_{\text{alt.}}$	x_{minor}	x_{major}
¬ rain	x_{minor}	$x_{\text{alt.}}$	x_{minor}	

5.6. The “Fallacies”

We conclude with a brief discussion concerning the fallacies of affirming the consequent and denying the antecedent. Only a relatively small number of experiments have studied the effects of Additional and Alternative premises on these arguments, as we mentioned earlier. However, the existing data suggest that the percentage of participants that endorse these fallacious inferences drops when an Alternative premise is present (as in AC Alternative below), but not when an Additional premise is present (see Fig. 1).

(16) **AC Additional:** If it was raining, Alicia got wet. If Alicia forgot her umbrella, Alicia got wet. Alicia got wet. *Therefore*, It was raining.

(17) **AC Alternative:** If it was raining, Alicia got wet. If she fell into a pond, Alicia got wet. Alicia got wet. *Therefore*, It was raining.

We note here that fewer participants endorse Affirming the Consequent in its Basic form than endorse Basic modus ponens (see Fig. 1 again). For this reason (and because of the paucity of data), our suggestions about these arguments are tentative.

Our framework predicts that neither of these inferences ought to be endorsed. Assuming that the context evolves in the ways our theory predicts, nothing rules out the possibilities in which ¬rain & wet from either scoreboard. What needs to be explained, then, is why there is any inclination to endorse these fallacious arguments and also why AC Additional is more tempting than AC Alternative. A first step in this explanation is the observation that people sometimes strengthen the Basic conditional so that it conveys its full converse:

(18) If it wasn’t raining, Alicia did not get wet.

Our hypothesis is this strengthening is invited, to a degree, by Additional premises and to a much lesser degree by Alternative premises.

Our development of this hypothesis here is purely theoretical. In Note 7, we mentioned that von Fintel (2001b) argues that, absent further contextual information, Gricean principles entail that the Basic conditional implicates only that *it is not the case*

that, no matter what, *Alicia got wet*—that is, that there is some salient antecedent *B* such that it is *not* the case that *if B, Alicia got wet*. That is well short of (18). Sometimes, however, a conditional implicates not only that some *B* exists such that *(if B)(D)* is false, but also that every salient *B* is such that *(if B)(D)* is false. In such cases, *A* is the only antecedent for which *(if A)(D)* is true, so that *A* is not only sufficient but also necessary for *D*. This strengthened implicature—effectively, the claim that *if* gets strengthened to *iff*—is called *conditional perfection* (Geis & Zwicky, 1971; Horn, 2000).¹⁰ As von Fintel highlights, many conditionals cannot get the conditional perfection implicature. For example, *Jane will be upset if you call her after midnight* does not implicate that she won't be upset under other circumstances. However, following Cornulier (1983), he also points out that the stronger implicature may arise when:

(G1) Conversational participants expect, given the utterance situation, that the speaker is trying to exhaust the range of conditions that suffice for the consequent of the conditional (von Fintel, p. 10, discussing Cornulier, 1983).

(G2) No other sufficient conditions (for [the consequent of the conditional]) are made salient within the discourse.

These two conditions can help us understand the pattern of suppression for the fallacies. While both (G1) and (G2) intuitively apply to Additional AC, (G2) seems violated in Alternative AC. Consider Additional AC. Suppose some participants assume that the experimenter is supposed to be listing all sufficient conditions for the consequent (*Alicia got wet*) and that no sufficient conditions, other than rain, are in play. If the perfection implicature is triggered, the final scoreboard for Additional AC is:

Additional AC (final)	remember		-remember	
	wet	¬ wet	wet	¬ wet
rain		X _{minor}		X _{major} , X _{additional}
¬ rain		X _{minor}	X _{perfection}	X _{additional} , X _{minor}

Note that our original hypothesis was that the state does not extend to include ¬rain & wet & remember possibilities on the ground that they are more remote than the other kinds of remember-possibilities.

However, the strengthening to perfection implicature might not be available in the context of an Alternative premise. This premise (e.g., *If she fell into a pond...*) offers a salient alternative that is sufficient for Alicia to get wet. Even participants who

initially accept (18) are unlikely to maintain it in the face of the Alternative conditional. In this case, possibilities exist in which pond & wet & \neg rain, and these suppress the inference.

Alternative AC (final)	pond		\neg pond	
	wet	\neg wet	wet	\neg wet
rain		χ_{minor} , χ_{alt}		χ_{major}
\neg rain		χ_{minor} , χ_{alt}		χ_{minor}

In sum, the contrast between Alternative premises (*If she fell into a pond...*) and Additional premises (*If she did not remember her umbrella...*) is that the latter does not introduce an independent sufficient condition and may allow participants to hold onto (18), vindicating the argument.

The above is merely an initial sketch of a full account of the fallacies. To complete it, we would need to provide empirical evidence that the limited distribution of the perfection implicature is as we suggest, and we ought to sharpen the notion of “sufficiency,” which we have left unanalyzed. We leave these tasks for separate work, because, if our approach is on the right track, an account of the fallacies is relatively independent from the core account of the suppression effect.

6. General discussion

Modus ponens is a paradigm case of deductive reasoning. Nearly all participants agree that the conclusion of modus ponens follows logically from its premises. Their accuracy in doing so rivals that of virtually any other item in cognitive psychologists’ repertoire of standard tasks. A participant who responds that q does *not* logically follow from *If p then q* and p is more likely to have confused the positions of the “yes” and “no” response keys than to have genuine doubts about the inference’s correctness. And yet many experiments, including our own, have shown that simply adding an extra premise is enough to cause many participants—sometimes the majority—to reverse their stance. This was the case, for instance, in Experiment 1 when we added a Contravening conditional to the Basic modus ponens premises.

What accounts for this odd combination of modus ponens’s acceptability and vulnerability? Our answer is that conditional sentences have special semantic and pragmatic features that make their interpretation sensitive to other sentences within an argument. On the semantic side, we claim that indicative conditionals are context-dependent modal constructions. A true indicative conditional is one in which the

conditional's consequent is true in all contextually relevant situations in which its antecedent is true. A conditional like *If it was raining, Alicia got wet* means that Alicia got wet in all raining situations that are consistent with the current presumptions of the speaker and hearer. On the pragmatic side, conditional sentences produce some presuppositions and implicatures of their own. One of these is that, as far as the current presuppositions go, the antecedent might be true. Appropriate use of *If it was raining, Alicia got wet* requires that the context be compatible with the possibility of rain (moreover, by Gricean reasoning, it should also be compatible with the opposite possibility: no rain). A second pragmatic effect of conditionals is that the context, after accepting a conditional, contains no situations in which the antecedent is true but the consequent is false. Although both rain and no rain are possibilities, Alicia getting wet accompanies rain.

These properties of conditionals can give rise to suppression because of the way reasoners update contexts while evaluating an argument's premises. In the case of Basic modus ponens, we are never forced to expand the salient state. The premises simply rule out possibilities. But for Contravening and Additional modus ponens, the extra conditional (e.g., *If she remembered her umbrella, Alicia did not get wet; If she forgot her umbrella, Alicia got wet*) requires the contextually determined state to expand to cover possibilities that at first we had ignored. In this expanded context, we introduce some worlds that would have been ruled out by the Basic conditionals (e.g., worlds in which it rains and Alicia does not get wet), undermining the argument's conclusion and producing suppression. Variations on these pragmatic adjustments can also explain suppression of modus tollens.

Although this is our main plot line, a number of unresolved subplot issues remain, some of which we've noted in passing. These issues divide into those concerning theories of reasoning and those concerning theories of conditionals. We consider these two sorts of questions in the rest of this article.

6.1. Relation to earlier theories

We believe that our proposal captures many of the insights of earlier approaches while avoiding some of their difficulties. In this section, we quickly review extant approaches to the suppression problem.

6.1.1. Mental model theories

Mental-model theory represents conditional sentences in a way that has much in common with the entries in a formal truth table in propositional logic (e.g., Barwise & Etchemendy, 2002). However, "each mental model represents a possibility. It is akin to a diagram in that its structure is analogous to the structure of the situation that it represents, unlike, say, the structure of logical forms..." (Johnson-Laird & Byrne, 2002, p. 647). The theory also allows some flexibility in its representation in order to accommodate limits on people's memory capacity and their background knowledge of the conditional's content. It is beyond the scope of this essay to give a detailed critique

of mental models (but see, e.g., Evans & Over, 2004; Rips, 2011; Stenning & Lambalgen, 2007). However, we highlight our differences in viewpoint to provide some perspective on our own proposal. We sketch the theory's account of suppression and vindication for modus ponens, but similar ideas extend to the other argument forms discussed earlier. The mental models approach to the suppression data was developed in Johnson-Laird and Byrne (1991) and Byrne et al. (1999), whose Experiment 1 was also designed to attack the probabilistic theories discussed below. Our exposition follows Byrne et al.

According to mental model theories (Johnson-Laird & Byrne, 1991, 2002), reasoners construct representations of conditionals, such as *If it was raining, Alicia got wet*, featuring explicit ways in which the conditional can be true. One mental model for this conditional would represent the situation in which it was raining and Alicia got wet, since in such a situation the conditional is true. If “rain” is a mental token standing for a situation in which it was raining and “wet” a token for one in which Alicia got wet, then

rain wet

would be a mental model that represents the combined situation in question. This representation should be taken literally as consisting of two mental tokens with no further internal structure, at least to the extent that internal structure is constructed from linguistic form (“models encode little or nothing of the linguistic form of the sentences on which they are based,” Johnson-Laird, 1983, p. 162). In some favorable cases, participants may create a fully explicit set of mental models that correspond to further situations in which the conditional is true, but typically reasoners construct only a subset of these models.

According to mental-model theory, a conditional's meaning does not determine the same mental model for every reasoning episode: A *modulation* process can make it the case that the very same sentence can end up associated with different mental models depending on the background context. Among such features are which possibilities have been introduced (e.g., the possibility that Alicia might have an umbrella in the case of the conditional *If it was raining, Alicia got wet*). In particular, the mental model associated with the major premise of modus ponens may vary according to which possibilities reasoners consider.

To account for suppression of modus ponens, mental-model theory proposes that an Additional premise (e.g., *If she forgot her umbrella, then Alicia got wet*), together with background knowledge, causes reasoners to represent the conditional premises in a new way. In the presence of Additional premises, reasoners' models are the same as those for a conditional in which the two antecedents are conjoined—in this case, *If it was raining and she forgot her umbrella, then Alicia got wet*. (Contravening premises, such as *If she remembered her umbrella, she did not get wet*, would presumably produce the same mental models.)

We share with mental-model theory the ideas that an account of the suppression effect ought to be grounded, in part, in the meaning of the conditional sentence and

that contextual factors can affect people’s willingness to draw inferences from conditionals. We also agree that people represent possibilities—worlds in which sets of statements are true—though we doubt they employ representations that are diagrammatic or “iconic” to do so (“in the sense that the structure of a model corresponds to the structure of what it represents,” Orenes & Johnson-Laird, 2012, p. 360). Our positive account differs because we take as our starting point one of the dominant views of the meaning of conditionals among linguists and philosophers of language, and this meaning is not that of Johnson-Laird and Byrne (2002). Instead, on our view, an indicative conditional’s meaning is fixed, as given by the **Strict** interpretation [i.e., *(if A)(B)* is true at w,i iff for every *A*-world v in i , *B* is true at v,i]. Suppression effects are the result of pragmatics principles (see Figure 3) altering the scope of the context (the scope of the state, i) in which the argument is evaluated.

This difference in initial starting point produces significant differences in our theory: Our account of suppression relies on general hypotheses about the interpretation of discourse. The mental-models approach has relatively few constraints on the way a conditional is interpreted. Pragmatic modulation allows the conditional “to refer to 10 distinct sets of possibilities out of the 16 a priori sets for binary connectives” (Johnson-Laird & Byrne, 2002, p. 660). For example, modulation allows a conditional such as *If it was raining, Alicia got wet* to be represented by (a) two models, in one of which the antecedent and consequent are both true and in the other of which the antecedent is true and the consequent false:

rain wet
rain -wet

or (b) one model in which both the antecedent and the consequent are false:

-rain -wet

or (c) four models representing all combinations of the truth of the antecedent and consequent:

rain wet
-rain -wet
rain -wet
-rain wet

and so on (see Johnson-Laird & Byrne, 2002, Table 4). Moreover, since the theory’s explanation of suppression depends on reinterpreting the major premise to include the antecedent of the Additional or Alternative premise (Byrne et al., 1999), the range of possible meanings for the conditional must expand to include sets of possibilities beyond the binary ones, with no obvious limits (cf. “The model theory postulates a mechanism of *modulation* in which the meanings of clauses, their referents, and knowledge, can transform the core meaning of connectives, including conditionals, into an indefinite number of other meanings,” Orenes & Johnson-Laird, 2012, p. 360, emphasis in original). According to the Principle of Pragmatic Modulation (Johnson-Laird & Byrne, 2002, p. 659):

The context of a conditional depends on general knowledge in long-term memory and knowledge of the specific circumstances of its utterance. This context is normally represented in explicit models. These models can modulate the core interpretation of a conditional, taking precedence over contradictory models. They can add information to models, prevent the construction of otherwise feasible models, and aid the process of constructing fully explicit models.

But although general knowledge and knowledge of context is important in understanding the suppression effect, mental-model theory has no clear account of when this information will add new models or modify old ones.

In our view, there is a reasonable alternative based on a uniform treatment of the conditional's meaning (e.g., **Strict**) and which assigns variations in its inference potential to shifts in common ground (or to effects of other sentence elements, such as modals). What govern these shifts, according to our theory, are the specific principles (e.g., **Context-set Possibility, Negative Possibility, Exclusion**), spelled out in Section 4. Of course, one could think of these principles as implementing "modulation" in that they take into account "long-term memory and knowledge of specific circumstances," but virtually any cognitive process would qualify as modulation in this sense. What's distinctive to the present approach, and absent in mental models, is the use of common ground within a formal pragmatic framework to capture the propositions that are relevant to inferences from conditional statements. There may be phenomena involving conditionals that call for modulation in the more general style of Johnson-Laird and Byrne (2002), but the suppression effect is not one of them.

6.1.2. Default theories

We've seen that a version of the mental-models theory (Johnson-Laird & Byrne, 1991), relies on the idea that an Additional or Alternative premise triggers a reinterpretation of the conditionals, relying on people's real-world knowledge. From *If it was raining, Alicia got wet* and *If Alicia forgot her umbrella, Alicia got wet*, reasoners produce a mental model that expresses *If it was raining and she forgot her umbrella, Alicia got wet*. An Alternative premise similarly produces a mental model but one that disjoins the Basic and Alternative antecedents. Direct evidence for reinterpretation comes from think-aloud protocols reported by Stenning and van Lambalgen (2007). Some participants explain their responses to these problems by paraphrasing the antecedents of the conditionals in conjunctive or disjunctive ways.

To account for this repackaging, Stenning and van Lambalgen propose that people represent the conditionals in a form that explicitly indicates default relations between the antecedent and consequent (they describe their conditional as "a special, noniterable non-truth-functional connective," p. 202). They show that people can arrive at conjunctive or disjunctive readings, similar to those of mental-models theory, by making two kinds of inferences: First, the conditionals themselves express relations

that hold only in the absence of exceptions or abnormalities, and second, the conditionals are governed by *closed-world assumptions* that only certain explicitly mentioned factors (or factors derived from inference) can be abnormalities.

As an example of Stenning and van Lambalgen's (2005, 2007) approach, consider Additional modus ponens. The theory represents the major premise, for example, *If it was raining, Alicia got wet*, as:

- (19) If it was raining and nothing abnormal happened, Alicia got wet.

Similarly, the Additional premise, *If Alicia forgot her umbrella, Alicia got wet*, would be represented as:

- (20) If Alicia forgot her umbrella and nothing abnormal' happened, Alicia got wet.

Here, *abnormal* in (19) indicates potential conditions that could prevent Alicia from getting wet in case of rain, and *abnormal'* in (20) indicates conditions (possibly different from those of *abnormal*) that could prevent her from getting wet in case she forgot her umbrella. People supplement these representations with two further conditionals about possible abnormalities, drawn from their background knowledge: (a) Remembering an umbrella counts as an abnormality with respect to getting wet in case of rain (*If Alicia did not forget her umbrella then it's abnormal*), and (b) not raining counts an abnormality' with respect to getting wet in the case of forgetting an umbrella (*If it was not raining then it's abnormal'*). These new conditionals have the same semantics as those associated with (19) and (20). Closed-world assumptions with respect to these conditionals then yield the result that *Alicia got wet if and only if (it was raining and she forgot her umbrella)*.

Our own explanation of suppression for modus ponens and tollens relies on the notion of typicality, and typicality may be viewed as closely connected to the notion of abnormality deployed by Stenning and van Lambalgen. A key difference is that, for us, typicality is not explicitly represented in the formal language and is only invoked in the account of how certain parameters receive their values. We view typicality as a pragmatic and cognitive constraint on the order in which people consider new possibilities that enlarge the set of those previously considered. It is worth noting, however, that the pragmatic elements in our account are essential parts of the interpretation of the conditional sentences, since the sentences' semantics depend on context. We are *not* proposing that people interpret an entire argument (e.g., Additional modus ponens) semantically and only then apply pragmatic corrections to the interpreted output. Instead, semantic and pragmatic processing is interleaved, with the pragmatics of one premise altering the set of relevant possibilities (the *i* set) that people use to interpret another. (Fig. 3 makes precise the sense in which this is true: Pragmatic adjustments associated with principles such as Composite Possibility affect

the context in which subsequent premises are evaluated.) The proposal is therefore easily consistent with recent evidence that effects of suppression occur just after participants finish reading the conclusion (Pijnacker, Geurts, van Lambalgen, Buitelaar, & Hagoort, 2011).

But why take this pragmatic-psychological route in preference to the semantic one? One reason concerns the closed-world assumption that the only abnormalities for a conditional are those explicitly mentioned. Consider, for example, the idea that not forgetting her umbrella is the one and only abnormality that could keep Alicia from getting wet if it rained. According to Stenning and van Lambalgen, this information comes from a list of abnormalities associated with the conditional *If it was raining, Alicia got wet*, with the list element not-forgetting-her-umbrella selected by world knowledge or the preceding text. However, people are able to understand a potentially infinite number of conditionals, and the call-off conditions for such statements seem entirely open-ended. Perhaps Alicia doesn't get wet in case of rain because she is indoors, is under a tree, is protected by rain gear, is beneath an awning, or is out of the path of the rain clouds, to name just a few of the more obvious ones.

Why, according to this approach, do people take into account remembering an umbrella but not conditions, such as wearing a raincoat, that might equally qualify as abnormalities? Stenning and van Lambalgen (2005, 2007) are no doubt right that mention of Alicia's forgetting the umbrella prompts reasoners to consider it an exception, but they also need the closed-world assumption (unmentioned or inferred factors can be ignored) to make it the only one. This restriction, however, seems to be an entirely pragmatic matter. Communicative relevance suggests that if a speaker knew of other factors that would have prevented Alicia from getting wet, s/he would have mentioned them. Not mentioning them implicates that they do not apply. This reasoning contrasts with the use of closed-world assumptions in the context of database search (e.g., if a customer database does not list Fred Smith, then Fred is not a customer), where the database's exhaustiveness justifies its application.

Stenning and van Lambalgen (2005, 2007) could agree that pragmatic factors are behind the decision to apply a closed-world assumption (or a similar restriction, such as exclusiveness) in the case of suppression. They sometimes speak of presupposition accommodation and other discourse factors (e.g., Stenning & van Lambalgen, 2007, p. 212). Our present point is that progress in understanding suppression depends on making explicit the reasoning that leads to decisions of this sort. This means being clear about the representations and processes that people employ at the pragmatic and contextual levels.

6.1.3 Pragmatic Theories

The accounts we just discussed assign some role to the pragmatics of conditional interpretation. Perhaps pragmatics can also prove more directly helpful in understanding suppression. Politzer (2005; Politzer & Braine, 1991) offers a pragmatic account of suppression that is similar in spirit to our explanation. We note that the

dividing line between semantics and pragmatics can shift, depending on a theory's logical framework. So we need to consider our semantics in conjunction with our pragmatic commitments.

Someone who states that *If it was raining, Alicia got wet* generally implicates that he or she doesn't know whether it was raining (where by *implicates* we mean that not knowing about the rain is a conversational implicature of the conditional, in the sense of Grice, 1989). Otherwise, it would be more appropriate to state the facts about the weather (and Alicia's resulting condition). According to the implicature approach, the same statement also implicates that raining was enough to ensure Alicia's getting wet. If we then learn *It was raining*, as the minor premise in Basic modus ponens, we cancel the implicature about whether it was raining but draw the conclusion of the argument. Now consider Additional modus ponens, with the extra premise *If she forgot her umbrella, then she got wet*. The Additional premise again implicates that the speaker doesn't know whether she forgot the umbrella, but also undermines (or *cancels*) the notion that raining was enough for Alicia to have gotten wet (if she remembered her umbrella, then she stayed dry). Thus, merely learning that it was raining no longer guarantees that she got wet, and modus ponens is suppressed.

We agree with the implicature theorists on one key point: A story about suppression of valid inferences ought to be based on a theory of the dynamics of conversation. However, the implicature theory lacks some crucial details. For one thing, the claim that rain is sufficient to get Alicia wet seems best regarded as part of the *content* of the conditional *If it was raining, Alicia got wet* rather than as an implicature. If it were a conversational implicature, it would be cancelable in the way that all such implicatures are (Grice, 1989). But it seems odd to say, *If it was raining, Alicia got wet; in fact, rain wasn't enough for her getting wet*. (Compare the ease with which a typical implicature can be canceled; e.g., *Some of Fred's children are in college; in fact, all of them are*.) In addition to this problem, the implicature story does not explain why we can restore endorsement of the conclusion in the case of Disjunctive modus ponens (see the example in Section 2.3): If the implicatures of the premises are canceled, why is the argument with the weaker conclusion vindicated? For example, suppose *If it was raining, Alicia got wet* and *If Alicia remembered her umbrella, she did not get wet* cancel each other's sufficiency implicature. If we then learn it was raining, we should not be able to conclude that *Either Alicia remembered her umbrella or Alicia got wet*. However, participants accept this inference on about 90% of trials (see Fig. 4).

Despite these problems, there is no denying the importance of pragmatic features in an account of the suppression effect. In fact, any explanation of the suppression phenomena will have to appeal to pragmatic facts like these. Mental models need something like them to account for why reasoners change their models for conditionals in the face of new possibilities. Probabilistic theories may need them to account for the change in conditional probabilities (see the quotation from Stevenson & Over, 1995, in the following section). Our framework, too, crucially relies

on a theory of pragmatics, though not solely on Grice's theory of conversational implicature.

6.1.4. *Undercutting approaches*

Another kind of explanations of suppression proposes that the added premises can undermine the relation between the premises and conclusion by lowering their probabilities. The most popular execution of this program is within a probabilistic theory of conditional reasoning (e.g., Chan & Chua, 1994; Stevenson & Over, 1995; Oaksford & Chater, 2007). Stevenson and Over (1995) offered a probabilistic account of suppression, drawing on work in philosophical logic by Adams (1975). Adams advanced a notion of probabilistic consequence, defined in terms of the technical notion of *uncertainty*—which for most propositions A is just $1-Pr(A)$. Adams's theory produces a notion of validity that is monotonic and validates modus ponens (Adams, 1998, p. 125).

Nonetheless, Stevenson and Over (1995) maintain that the Adams-style account can handle the suppression of modus ponens. Their crucial idea is that the conclusion of a valid argument may well have high uncertainty (i.e., low probability). On their view, participants do not respond to the validity of the argument, but rather to the degree of uncertainty of the conclusion.

If we get more information, perhaps only pragmatically, leading us to doubt the major premise, then we could be led to express doubt about the conclusion. [Participants] might “suppress” the conclusion by not inferring it at all. For though its uncertainty cannot be greater than the sum of the uncertainties of the premises, that is not now saying very much (p. 616).

In support of their view, Stevenson and Over explicitly manipulated the degree of uncertainties of the premises and of the conclusion in Additional modus ponens.

Oaksford and Chater (2007) also offer a probabilistic model that attempts to explain the degree of acceptance of various inferences as a function of certain probabilistic parameters. The inspiration of this model is similar to that of Stevenson and Over (1995): Oaksford and Chater also claim that participants' judgments about the acceptability of various inferences should be read as reflecting particular probabilistic states. They differ, however, on the type of state: While Stevenson and Over are concerned with the unconditional probability of the conclusion, Oaksford and Chater assume that the degree of acceptability of the inference is predicted by $Pr(\textit{conclusion} \mid \textit{minor premise})$.

We believe that this style of probabilistic analysis gets part of the story about suppression right. Centrally, we agree with the idea that participants do not necessarily answer the logical question: *Is the argument valid?* Furthermore, a probabilistic framework may have to be invoked once we consider pieces of reasoning in which the

premises are not assumed but are themselves treated as uncertain (Stevenson & Over, 2001).

With Stenning and van Lambalgen (2007, pp. 214-216), we are skeptical, however, that it is the complete story. For one thing, it is unclear how reasoners arrive at the salient probability models. Probabilistic theories offer no update rule that has the effect that entertaining a certain possibility will make a given conditional proposition more or less probable. For example, if the probability that Alicia got wet, given that it rained, is .85, how does the probability of this conditional change when we learn that she had her umbrella? This is why Stevenson and Over's explanation involves pragmatic processes. But if we go the pragmatic route, we must clarify what those processes are. On our view, pragmatic processes are enough to explain suppression, quite independently of the probabilistic machinery. (See, also, Jones & Love, 2011, and Marcus & Davis, 2013, for general critiques of probabilistic theories' lack of psychological specificity.)

Another difficulty for these explanations is that they predict that endorsement rates for *Basic* modus ponens should vary according to the probabilities that people attach to the premises. However, Basic modus ponens is usually resistant to this variation, and this creates a puzzle for probability theories. As Figs. 1 and 4 illustrate, people endorse Basic modus ponens at a rate greater than 90% even for the same major premises that figure in Contravening or Additional modus ponens arguments in the same experiments. The adult participants in these studies certainly know from prior experience about factors that can disrupt the sufficiency of the major premises (e.g., factors that could keep Alicia from getting wet if it rained). These factors, however, do not ordinarily cause them to doubt Basic modus ponens (see the following subsection for potential exceptions). Similarly, people surely assign higher probabilities to the major premise of the argument in (21) than the one in (22):

(21) If car X10 runs into a brick wall, it will stop. Car X10 runs into a brick wall.
Therefore, Car X10 will stop.

(22) If car X10 runs into a brick wall, it will speed up. Car X10 runs into a brick wall.
Therefore, Car X10 will speed up.

The greater uncertainty of (22) [vs. (21)] should mean that people would be less likely to draw the conclusion of (22) than the one of (21), by the very same logic that predicts lower endorsement of Additional modus ponens than Basic modus ponens. However, no such difference appears when participants are under instructions to say whether the conclusion necessarily follows from the premises (Rips, 2001). One could argue that those instructions are irrelevant to way people reason in naturalistic situations, but this does not explain the data.

To handle such effects, Oaksford and Chater (2007, p. 128) "allow the model to always capture the modus ponens inference and then fit the model for just DA, AC,

and MT.” But the account offered here seems more explanatory. Indicative conditionals come along with a requirement of exclusion: The context excludes worlds in which the antecedent is true and the consequent false. This supports Basic modus ponens, but the effects of the requirement can be overturned by explicit Additional premises.

6.1.5. *Implicit suppression*

We need to consider one final source of evidence about conditional inferences since this evidence bears on the theories just discussed, especially the probabilistic theory. People do not usually suppress modus ponens unless an explicit Additional or Contravening premise appears in the argument. However, exceptions occur in some experiments that have varied the conditionals' content. These studies have collected either: (a) estimates of the number of contravening situations that people can bring to mind for a conditional major premise (Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003), (b) ratings of the sufficiency of the antecedent for the consequent (Thompson, 1994), (c) ratings of the truth of the conditional (George, 1995), or (d) estimates of the frequency of exceptions (antecedent-and-not-consequent cases; Geiger & Oberauer, 2007). These factors all tend to suppress Basic modus ponens. The larger the numbers of contravening situations or exceptions and the lower the ratings of truth or sufficiency, the less willing participants are to endorse the argument. For example, Cummins et al. (1991) produced suppression with arguments such as: *If I eat candy often, then I have cavities; I eat candy often; therefore, I have cavities*. People know from their own experience that contravening factors, such as fluoride treatments, can disrupt the causal connection between eating candy and having cavities, and this knowledge produces lower agreement with the argument. We will refer to effects like these as *implicit suppression* to contrast it with the explicit suppression that depends on Contravening or Additional premises (see Byrne, 2005, p. 112).

Our own experiments used some of the same conditionals that appeared in these earlier studies but without yielding suppression. For example, four of the conditionals in the Basic modus ponens arguments of Experiment 1 came from Cummins (1995) and were among those normed as having many contravening factors (e.g., *If Jenny turned on the air conditioner, then she felt cool*). In our experiment, however, participants endorsed Basic modus ponens with these conditionals on 93-100% of trials. Why no implicit suppression for these items? One factor that could explain the difference between studies is a change in instructions and response requirements. Participants in Cummins (1995) responded by giving a rating on a scale that varied from “very sure that I *cannot* draw this conclusion” to “very sure that I *can* draw this conclusion.” In addition, participants were *not* told to assume the truth of the premises or to decide whether the conclusion logically followed. Instead, “the subjects were encouraged to reason as they would in everyday circumstances” (p. 652). Similarly, de Neys et al. (2003) found implicit suppression effects when participants used a rating scale like Cummins's and were told “to evaluate the conclusion by criteria they personally judged to be relevant” (p. 585). Geiger and Oberauer (2007) also

obtained suppression effects due to the frequency of counterexamples to the conditionals, but using rating scales rather than the usual follows/doesn't follow decision.

Response scales that measure relative certainty encourage participants to treat the problems as a matter of probabilistic reasoning (Evans, Handley, Neilens, & Over, 2010; Markovits, Forgues, & Brunet, 2010; Markovits & Handley, 2005). So do instructions that fail to ask participants whether the premises entail the conclusion (Evans et al., 2010). We know of three experiments, however, that have found implicit suppression effects with standard deduction instructions and response formats. George (1995, Experiment 3) used conditionals "with controversial topics, in order to elicit possible doubts or disagreements" (p. 96) and found 71% endorsement of Basic modus ponens. Similarly, Thompson (1994, Experiment 1) found 84% endorsement with conditionals, such as *If weather conditions are bad, then the plane will crash*, that had been judged as having antecedents insufficient to produce their consequents. Finally, Evans et al. (2010) divided their participants into high- and low-ability groups based on a general intelligence test and obtained 75% endorsements among the low-ability group for conditionals rated as low in believability (e.g., *If fast food is taxed then childhood obesity will increase*). Although these percentages are not as low as those produced by Additional premises, they are lower than those for Basic modus ponens in standard suppression experiments (see Fig. 1). The reason for this intermediate effect is unclear, as is the reason why it turns up in some experiments but not others. However, as George points out, some participants fail to follow instructions to assume the premises are true, and this tendency may be exacerbated by controversial or uncertain conditionals and by low ability.

Rating scales and instructions to evaluate the arguments informally may tempt participants to assess the perceived strength of the relation between the antecedent and consequent: the degree of confidence that should be placed in the consequent (i.e., the conclusion of the modus ponens argument) given the antecedent (the minor premise). Probabilistic theories of conditional reasoning (e.g., Oaksford & Chater, 2007; Stevenson & Over, 1995) provide models for this procedure. Our claim, therefore, is not that participants never adopt such a strategy. What we doubt is that this is the only method people use.

Studies of implicit suppression sometimes justify their experimental procedures by asserting that they come closer to normal modes of reasoning. Cummins et al. (1991), for example, defend not instructing participants to assume the truth of the premises on the grounds that "we were interested in observing human reasoning performance as it typically occurs, and not in instructing subjects in deductive logic" (p. 278). However, reasoning "as it typically occurs" involves adopting the presuppositions that go along with the assertion of a conditional, and these include the idea that the relevant context has no situations in which a conditional's antecedent is true but its consequent is false. Settings that emphasize the conditional's uncertainty

may cancel this presupposition, but that does not make the presupposition an unnatural one.

6.2. *Conditional semantics and pragmatics*

Our investigation was guided by the idea that suppression can be explained by established ideas about context dynamics. However, it is also worth indicating the possibility of the development of formal semantics from results in the psychology of reasoning: In order to apply the established ideas to the suppression data, we had to introduce tweaks on (and specifications of) the standard semantic apparatus. If data about suppression of modus ponens and modus tollens have the same standing as the data that are used in philosophy of language and semantics, our analysis can be viewed as putting forward new constraints on semantic theories of conditionals, as well as new ideas about how to satisfy them.

For instance, earlier on, we have considered the argument with premises:

(a) If it was raining, then Alicia got wet. (b) If it was raining and Alicia remembered her umbrella, Alicia did not get wet. (c) It was raining and Alicia remembered her umbrella.

The literature on Sobel Sequences, such as the sequence (a)-(b) in this example, invokes two broad kinds of explanation for why these pairs are compatible. The first type involves introducing new notions of logical consequence (von Fintel, 2001a; Gillies, 2007); this is the relation that the premises have to the conclusion of a (semantically) valid argument. The other type of explanation involves general principles concerning the dynamics of conversation (Moss, 2012; Williams, 2008).

It seems to us that, if we are right in thinking that Additional modus ponens and Contravening modus ponens form a unified kind, the conversational explanation has an advantage. This is because the revised notion of consequence is only sensitive to what is built into the meaning of an expression (possibly including its presuppositions). It does not model updates that are due to Gricean facts, and we invoked these updates in our explanation of why Additional inferences are suppressed. By contrast, on the conversational explanation, the evaluation of the argument is treated more like the interpretation of a story: Successive updates are entertained that might be due to presupposition, meaning, or implicature.

To be anything more than a programmatic sketch, this argument needs a lot of refinement, and we defer this task to future work. However, even at this very coarse level of generality, it is clear that data from the literature on the suppression effect can further our understanding of the semantics and pragmatics of conditionals. The benefits of interaction between the psychology of conditional reasoning and the semantics of conditionals go both ways.

Acknowledgments

We are grateful to two anonymous referees for *Cognitive Science*, Susan Brennan, Ruth Byrne, Jacob Dink, Catarina Dutilh Novaes, Kai von Fintel, John Glines, Jacqueline Herschey, Samuel Johnson, Joshua Knobe, Nicholas Leonard, David Over, and Keith Stenning for help on preliminary versions of this article. We also thank the audience at the 2013 Philosophy of Language and Linguistics Workshop at the Interuniversity Center in Dubrovnik, where it was presented.

References

- Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: Reidel.
- Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: CSLI Publications.
- Barwise, J., & Etchemendy, J. (2002). *Language, proof and logic*. Stanford, CA: CSLI Publications.
- Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford, UK: Oxford University Press.
- Boghossian, P. (2000). Knowledge of logic. In P. Boghossian & C. Peacocke (Eds.), *New essays on the a priori* (pp. 229-254). Oxford, UK: Oxford University Press.
- Braine, M. D. S., Reiser, B. J., & Rumin, B. (1998). Evidence for the theory: Predicting the difficulty of propositional logic inference problems. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic*. (pp. 91-144). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31, 61-83. doi: 10.1016/0010-0277(89)90018-8
- Byrne, R. M. J. (2005). *The rational imagination*. Cambridge, MA: MIT Press.
- Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, 40, 347-373. doi: 10.1006/jmla.1998.2622
- Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, 53, 217-238. doi: 10.1016/0010-0277(94)90049-3
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Cornulier, B. (1983). 'If' and the presumption of exhaustivity. *Journal of Pragmatics*, 7, 247-249.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, 23, 646-658. doi: 10.3758/bf03197265
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, 19, 274-282. doi: 10.3758/bf03211151
- De Neys, W., Schaeken, W., & D'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*, 31, 581-595. doi: 10.3758/bf03196099
- Dieussaert, K., Schaeken, W., Schroyens, W., & d'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking & Reasoning*, 6, 125-160. doi: 10.1080/135467800402820
- Dummett, M. (1991). *The logical basis of metaphysics*. Cambridge, MA: Harvard University Press.
- Edgington, D. (1995). On conditionals. *Mind*, 104, 235-329. doi: 10.1093/mind/104.414.235
- Espino, O., Santamaria, C., & Byrne, R. M. J. (2009). People think about what is true for conditionals, not what is false: Only true possibilities prime the comprehension

- of "if". *The Quarterly Journal of Experimental Psychology*, 62, 1072-1078. doi: 10.1080/17470210802602375
- Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *The Quarterly Journal of Experimental Psychology*, 63, 892-909. doi: 10.1080/17470210903111821
- Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Erlbaum.
- Evans, J. S. B. T., & Over, D. E. (2004). *Ifs*. Oxford: Oxford University Press.
- von Fintel, K. (2001a). Counterfactuals in a dynamic context. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press.
- von Fintel, K. (2001b). Conditional Strengthening. A Case Study in Implicature. Unpublished Manuscript, available at: <http://web.mit.edu/fintel/fintel-2001-condstrength.pdf>
- Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition*, 35, 2060-2074. doi: 10.3758/bf03192938
- Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry*, 2, 561-566. doi: 10.2307/4177664
- George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology*, 86, 93-111. doi: 10.1111/j.2044-8295.1995.tb02548.x
- Gillies, A. S. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy*, 30, 329-360. doi: 10.1007/s10988-007-9018-6
- Gillies, A. S. (2009). On truth-conditions for if (but not quite only if). *Philosophical Review*, 118, 325-349. doi: 10.1215/00318108-2009-002
- Gillies, A. S. (2010). Iffiness. *Semantics & Pragmatics*, 3, 1-42. doi:10.3765/sp.3.4
- Giroto, V., Mazzocco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition*, 63, 1-28. doi: 10.1016/s0010-0277(96)00792-5
- Grice, P. (1989). Logic and conversation. In *Studies in the way of words* (pp. 1-143). Cambridge, MA: Harvard University Press.
- Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 805-812. doi: 10.1037/a0018784
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, 118, 248-271. doi: 10.1037/0033-2909.118.2.248
- Horn, L. R. (2000). From if to iff: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics*, 32, 289-326. doi: 10.1016/S0378-2166(99)00053-3
- Ippolito, M. (2003) Presupposition and implicature in counterfactuals. *Natural Language Semantics*, 11, 145-186. doi: 10.1023/a:1024411924818

- Jackson, F. (1991) *Conditionals*. Oxford, UK: Oxford University Press.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646-678. doi: 10.1037/0033-295x.109.4.646
- Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34, 169-188. doi: 10.1017/s0140525x10003134
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454. doi: 10.1016/0010-0285(72)90016-3
- Kolodny, N., & MacFarlane, J. (2010). Ifs and oughts. *Journal of Philosophy*, 107, 115-143.
- Knobe, J., & Szabo, Z.G. (2013). Modals with a taste of the deontic. *Semantics & Pragmatics*, 6, 1-42. doi: 10.3765/sp.6.1
- Kratzer, A. (1981). The notional category of modality. In H. J. Eikmeyer & H. Rieser (Eds.), *Words, worlds, and contexts* (pp. 38-74). Berlin: de Gruyter.
- Kratzer, A. (1991). Conditionals. In A. von Stechow & D. Wunderlich (Eds.), *Semantics: An international handbook of contemporary research* (pp. 651-656). Berlin: de Gruyter.
- Lewis, C.I. (1918). *A survey of symbolic logic*. Berkeley: University of California Press. Reprinted by Dover Publications (New York), 1960, with the omission of Chapters 5-6.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8, 339-359. doi: 10.1007/978-3-642-67458-7_12
- Lycan, W. (1993), 'MPP, RIP'. *Philosophical Perspectives*, 7, 411-428. doi: 10.2307/2214132
- Lycan, W. (2001). *Real conditionals*. Oxford: Oxford University Press.
- Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, 24, 2351-2360. doi: 10.1177/0956797613495418
- McGee, V. (1985). A counterexample to modus-ponens. *Journal of Philosophy*, 82, 462-471. doi: 10.2307/2026276
- Markovits, H., Forgues, L. H., & Brunet, M.-L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition*, 38, 485-492. doi: 10.3758/mc.38.4.485
- Markovits, H., & Handley, S. (2005). Is inferential reasoning just probabilistic reasoning in disguise? *Memory & Cognition*, 33, 1315-1323. doi: 10.3758/bf03193231

- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Moss, S. (2012). On the pragmatics of counterfactuals. *Nous*, 46, 561-586. doi: 10.1111/j.1468-0068.2010.00798.x
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.
- Orenes, I., & Johnson-Laird, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language*, 27, 357-377. doi: 10.1111/j.1468-0017.2012.01448.x
- Parsons, L. M., & Osherson, D. (2001). New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cerebral Cortex*, 11, 954-965. doi: 10.1093/cercor/11.10.954
- Partee, B. H. (1989). Possible worlds in model-theoretic semantics: A linguistic perspective. In S. Allén (Ed.), *Possible worlds in humanities, arts, and sciences* (pp. 93-123). Berlin: de Gruyter.
- Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2011). Reasoning with exceptions: An event-related brain potentials study. *Journal of Cognitive Neuroscience*, 23, 471-480. doi: 10.1162/jocn.2009.21360
- Politzer, G. (2005). Uncertainty and the suppression of inferences. *Thinking & Reasoning*, 11, 5-33. doi: 10.1080/13546780442000088
- Politzer, G., & Braine, M. D. (1991). Responses to inconsistent premisses cannot count as suppression of valid inferences. *Cognition*, 38, 103-108. doi: 10.1016/0010-0277(91)90023-w
- Ramsey, F. P. (1965). General propositions and causality. In R. B. Braithwaite (Ed.), *Foundations of mathematics and other logical essays* (pp. 237-255). Totowa, NJ: Littlefield, Adams. Original article written in 1929
- Rips, L. J. (1994). *Psychology of proof*. Cambridge, MA: MIT Press.
- Rips, L. J. (2011). *Lines of thought: Central concepts in cognitive psychology*. Oxford, UK: Oxford University Press.
- Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, 12, 129-134. doi: 10.1111/1467-9280.00322
- Rips, L. J., & Marcus, S. L. (1977). Suppositions and the analysis of conditional sentences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 185-220). Hillsdale, NJ: Erlbaum.
- Rotello, C. M., & Heit, E. (2009). Modeling the effects of argument length and validity on inductive and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1317-1330. doi: 10.1037/a0016648
- Rumain, B., Connell, J., & Braine, M. D. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: *If* is not the biconditional. *Developmental Psychology*, 19, 471-481. doi: 10.1037/0012-1649.19.4.471
- Schechter, J., & Enoch, D. (2006). Meaning and justification: The case of Modus Ponens. *Nous*, 40, 687-715. doi: 10.1111/j.1468-0068.2006.00629.x

- Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 140-149. doi: 10.1037/0278-7393.29.1.140
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Erlbaum.
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 31-95. doi: 10.1016/0010-0277(95)00666-m
- Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory*. Oxford, UK: Blackwell.
- Stalnaker, R. (1970). Pragmatics. *Synthese*, 22, 272-289. doi: 10.1007/BF00413603
- Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, 2, 447-457. doi: 10.1007/BF00262951
- Stalnaker, R. (1975). Indicative conditionals. *Philosophia*, 5, 269-286. doi: 10.1007/Bf02379021
- Stalnaker, R. (1978). Assertion. *Syntax and Semantics*, 9, 315-332.
- Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
- Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, 25, 701-721.
- Stalnaker, R. (2011). Conditional propositions and conditional assertions. In A. Egan & B. Weatherson (Eds.), *Epistemic modality* (pp. 227-248). Oxford, UK: Oxford University Press.
- Stenning, K. & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, 29, 919-960.
- Stenning, K., & van Lambalgen, M. (2007). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.
- Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 48A, 613-643. doi: 10.1080/14640749508401408
- Stevenson, R. J., & Over, D. E. (2001). Reasoning from uncertain premises: Effects of expertise and conversational context. *Thinking and Reasoning*, 7, 367-390. Doi: 10.1080/13546780143000080
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, 22, 742-758. doi: 10.3758/bf03209259
- Weatherson, B. (2009). Conditionals and indexical relativism. *Synthese*, 166, 333-357. doi: 10.1007/s11229-007-9283-5
- Willer, M. (2013a). Dynamics of epistemic modality. *Philosophical Review*, 122, 45-92.
- Willer, M. (2013b). Indicative scorekeeping, *Proceedings of the 19th Amsterdam Colloquium*: 249-256, 2013.
- Williams, J. R. G. (2008). Conversation and conditionals. *Philosophical Studies*, 138, 211-223. doi: 10.1007/s11098-006-9032-5
- Williamson, T. (2007). *The philosophy of philosophy*. Oxford, UK: Blackwell.

Yalcin, S. (2007). Epistemic modals. *Mind*, 116, 983-1026. doi: 10.1093/Mind/Fzm983

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

Notes

¹ None of this is to imply that modus ponens is without its challengers (in the philosophical literature, see McGee, 1985; Lycan, 1993, 2001; MacFarlane & Kolodny, 2010). Likewise, philosophers have questioned the idea that modus ponens is part of the conditional's meaning (Williamson, 2007). See also Schechter and Enoch (2006) for an argument to the effect that even if modus ponens is constitutive of the meaning of *if*, this is not sufficient to explain why we are epistemically justified in making modus ponens inferences.

² See von Fintel (2001a) and Gillies (2007) for discussions of this strategy with respect to counterfactual conditionals, and Williams (2008) for a similarly inspired, but formally very different, strategy with respect to indicatives.

³ This claim must be qualified: The contextually determined range of possibilities need not include the actual world. When it does not, it will not be literally true that **Strict** is intermediate in strength, as there might be conditionals that are true according to **Strict**, but false according to **Material**.

⁴ The difference between **Strict** and **Variably Strict** shows up as a difference in the verdicts of logical consistency. Consider the pair:

- If he took the exam, he passed it.
- If he took the exam and was caught cheating, he did not pass it.

According to **Variably Strict**, these two conditionals can both be true relative to a single pair $\langle w, R \rangle$ consisting of a world and a similarity relation. To evaluate the first we reach out to the most similar world v in which he took the exam and evaluate whether in that world the exam was passed. To evaluate the second, we reach out to a world z in which he took the exam and was caught cheating. Worlds v and z need not be identical: In fact, ordinarily, we may suppose that the most similar world in which the exam was taken is one in which no cheating occurs.

According to **Strict**, however, there is no single pair $\langle i, w \rangle$ that makes both sentences true. Nonetheless, they may still be, in a sense, jointly acceptable. As Stalnaker (1984) points out in the passage immediately following the earlier quotation (p. 125), someone who defends **Strict** may propose that, when we accept these statements in sequence, we are subtly, but systematically, switching contexts. Our approach to suppression posits a similar sort of context switch.

⁵ Evans and Over (2004) object to the strategy of Rips and Marcus (1977) on the grounds that it requires reasoners to have unbounded computational capacity, in order to grasp the infinite totality of possible worlds and to manipulate in their reasoning abstract objects that have the standard characteristics of possible worlds. Whether or not this objection applies to Rips and Marcus (1977), it will be clear that it doesn't apply to the present account. All that we need in order to characterize the effect of conditional reasoning is the idea that the total space of possibilities can be partitioned in finitely many alternatives and that at various stages in a conversation these alternatives can be ruled out as incompatible or ruled in as compatible with the current state.

⁶ The Composite Possibility principle is consistent with evidence that a conditional $(if A)(C)$ does not call to mind possibilities in which A is true and C is false (see Espino, Santamaria, & Byrne, 2009), since these possibilities are eliminated by Exclusion, as we are about to see. What it does stipulate is that when context changes to include new salient possibilities (e.g., D), people will consider these new possibilities in conjunction with A . Some of these novel possibilities may include ones (e.g., $A \& \neg C \& D$) that might have been eliminated had they been considered initially, but now seem reasonable in light of the new considerations. Our account of Contravening and Additional modus ponens will turn on this idea, and we know of no empirical results that contradict it.

⁷ Von Fintel (2001b) persuasively argues that a conditional of the form $(if A)(C)$ normally implicates that there are salient possibilities B such that $(if B)(C)$ is false. In these conditions, in other words, $(if A)(C)$ implicates that C does not hold no matter what. In some situations, however, this implicature can be strengthened, as we discuss in Section 5.6.

⁸ In a recent paper, Stalnaker (2011) has proposed something very similar to Exclusion in characterizing the effect of conditionals in theories that postulate the existence of a distinctive kind of speech act of *conditional assertion*.

First, one adds the content of the antecedent, temporarily, to the context; that is, one sets aside the possibilities in the context set in which the supposition is false.[...] Then the content of the consequent is treated like the content of a categorical assertion: one eliminates, from this temporary or derived context those possible situations that are incompatible with the content of the consequent. Finally, one adds back the possibilities that one had set aside. (Stalnaker 2011, p. 239)

We do not want to commit to a *conditional assertion* theory here, but we think that both Stalnaker's quotation and our principle of Exclusion are intuitive descriptions of the pragmatic effect of an utterance of a bare conditional.

⁹ Given our emphasis on connections between indicative and counterfactual conditionals, we should also mention that the asymmetry between modus ponens and modus tollens disappears in the case of counterfactual conditionals. Though we have no space for an adequate treatment of the issue, we note that this is not incompatible with our view. There is a long-standing consensus in the semantics literature that counterfactual conditionals carry an implicature to the effect that their antecedents are false (this point has a long history, but for a recent extensive treatment see Ippolito, 2003). If so, our *default* prediction ought to be that the endorsement rates for modus tollens involving counterfactuals should be higher.

¹⁰ Earlier research on conditional reasoning has used conditional perfection to explain why people sometimes accept affirming the consequent and denying the antecedent (e.g., Romain et al., 1983). We differ from this literature in that we follow Cornulier and von Fintel in holding that it has very special licensing conditions, which is what helps us to distinguish Additional and Alternative versions of the fallacies.