The crucial role of compositional semantics in the study of reasoning*

Daniel Lassiter[1]

School of Philosophy, Psychology, and Language Sciences
University of Edinburgh
dan.lassiter@ed.ac.uk

Abstract. Reasoning work in both psychology and computer science is often conceived in terms of drawing inferences over strings. This approach works reasonably well in some cases, but it is sometimes dramatically misleading due to the pervasive presence in natural language of interpretive factors that are not reflected transparently in surface form. The problem includes not only the well-known issues of ambiguity, polysemy, and sensitivity to pragmatic context, but also a much less-studied factor: compositional semantic interactions where one expression in a sentence influences the interpretation of another. In two case studies, I show that careful attention to natural language semantics is crucial in understanding valid and invalid patterns of reasoning, and in avoiding fallacies.

Keywords: Psychology of reasoning · Computational reasoning · Compositional semantics · Validity.

Linguistics is extremely important in psychological and computational work on reasoning. This has been recognized for some time, especially in the domains of pragmatics and lexical semantics. But compositional semantics—the study of how the form of sentences relates to their semantic interpretation—has not been given sufficient attention, especially in psychology. This paper draws attention to two ways that attention to semantics can clarify core issues in the study of reasoning, and help us avoid fallacious arguments.

1 Counter-examples to bedrock reasoning principles

Psychological research on reasoning has traditionally been concerned with the validity or invalidity of various principles of inference, mostly drawn from philosophical logic. Conditional reasoning is by far the most intensively investigated

* I'm very grateful to the organizers of the conference "Human and Artificial Rationalities" for their kind invitation to give a plenary talk. Thanks also to David Over for very helpful discussion, to three anonymous reviewers for this volume, and to Jean Baratgin, Maxime Boulier, and Baptiste Jacquet, whose experimental collaboration inspired a major portion of the thoughts below.

domain, and Modus Ponens (MP) is surely the best-supported non-trivial principle of conditional reasoning, being almost unanimously endorsed in numerous studies [6]. Modus Tollens (MT) is the ugly twin of MP: while participants endorse it somewhat less enthusiastically, most researchers consider MT to be valid nonetheless, attributing reduced endorsement to the psychological difficulty of applying the principle.

(1) Modus ponens:
   a. If A, C.
   b. A.
   c. Therefore, C.

(2) Modus tollens:
   a. If A, C.
   b. Not C.
   c. Therefore, not A.

MP and MT sure do look like valid rules of inference. When we fill in the variables with simple English sentences, the reasoning seems rock-solid.

(3) a. If it's raining, Mike is carrying an umbrella.
   b. It's raining.
   c. Therefore, Mike is carrying an umbrella.

(4) a. If it's raining, Mike is carrying an umbrella.
   b. Mike is not carrying an umbrella.
   c. Therefore, it's not raining.

Constructive Dilemma (CD) is a more complex principle, but one that is widely considered valid. Here it is in the abstract:

(5) a. If A, C.
   b. If B, D.
   c. A or B.
   d. Therefore, C or D.

Again, filling in the variables with simple English sentences gives us what looks like impeccable reasoning:

(6) a. If it's raining, Mike is carrying an umbrella.
   b. If it's sunny, he's wearing sunglasses.
   c. It's either raining or it's sunny.
   d. Therefore, either Mike is carrying an umbrella or he's wearing sunglasses.

I have bad news: all three principles admit of obvious counter-examples. Here is one.

(7) a. If it's raining, Mike always carries an umbrella.
   b. It's raining.
   c. Therefore, Mike always carries an umbrella.

This seems to be an instance of the MP template above in (1), substituting *It's raining* for A and *Mike always carries an umbrella* for C. It's also a terrible argument. If we assume that the generalization in (7a) holds, and that it is raining right now, we still cannot conclude (7c) unless we have some *additional* information—namely, that it always rains. So this is a counter-example to MP.

Here is a similar counter-example to MT:

(8) a. If it's raining, Mike always carries an umbrella.
 b. It's not the case that Mike always carries an umbrella.
 c. Therefore, it's not raining.

Suppose that Mike carries an umbrella always when it rains, and never carries one when it's sunny. Furthermore, we live in a place where it rains on half of the days, and is sunny on the others. Then both (8a) and (8b) are true. But this knowledge doesn't allow us to conclude anything about the *current* state of the weather in (8c)! What has gone wrong?

For good measure, here's the matched counter-example to CD.

(9) a. If it's raining, Mike always carries an umbrella.
 b. If it's sunny, Mike always wears sunglasses.
 c. It's either raining or it's sunny.
 d. Therefore, either Mike always carries an umbrella, or he always wears sunglasses.

Since one counter-example is enough to show that a rule of inference is not valid, this would seem to imply that MP, MT, and CD are not valid! Have we just overturned the intuitions of generations of philosophers and logicians, and the hard-won results of decades of careful psychological research? The answer is, as we'll see below, is "It depends". There is definitely something fishy going on in the counter-examples, but we need help from research in compositional semantics to see what it is and how we can respond.[1]

2 Foibles of string-based reasoning

Semantic theory sheds important light on the puzzle just posed. Lewis [23] pointed out that conditional antecedents interact compositionally with adverbs of quantification (AQs) like *always, frequently, seldom*, etc. When an AQ appears in the immediate scope of an *if*-clause, the *if*-clause is usually interpreted as providing a restriction on the quantificational domain of the AQ. As a result, the item *always* is interpreted in a special way in (the most prominent reading of) the major premise of the arguments above.

[1] Yalcin [32] gives a counter-example to MT similar to (8), but does not point out that the problem affects MP equally. I will discuss Yalcin's argument further below. McGee [27] famously gives a purported counter-example to MP involving nested conditionals, but also contends that MP is valid in other contexts. In any case, many have found McGee's counter-example less than fully convincing (e.g., [2, 28]).

(10) If it's raining, Mike always carries an umbrella.

a. **Wrong**: If it's raining (now), Mike carries an umbrella *in all relevant situations*.

b. **Right**: *In all relevant situations in which it's raining*, Mike carries an umbrella.

All of the purported counter-examples given above rely on this kind of shift in the domain of an AQ. When *always* occurs in the consequent of a conditional—e.g., the major premise of our counter-examples to MP and MT—it quantifies over a restricted domain of situations that are contextually relevant *and* satisfy the conditional antecedent. When *always* occurs unembedded, it quantifies over a larger domain: the situations that are contextually relevant, full stop. The schematic interpretations given below make clear why these arguments are not convincing. Let R be a variable standing for the contextually relevant situations. Lewis' analysis implies that the purported MP counterexample (11) is interpreted along the lines of (12), and the MT counterexample (13) as in (14).[2]

(11) a. If A, always C.

b. A.

c. Therefore, always C.

(12) a. $ALL[R \cap A][C]$.

b. A (at speech time).

c. Therefore, $ALL[R][C]$.

(13) a. If A, always C.

b. Not always C.

c. $\therefore$ Not A.

(14) a. $ALL[R \cap A][C]$.

b. $\neg ALL[R][C]$.

c. $\therefore \neg A$ (at speech time).

In the MP counter-example (11), the item *always* is interpreted differently in the major premise and the conclusion—as the schematic interpretation in (12) makes clear. Specifically, the restriction of the *ALL*-quantifier that *always* denotes is different in the two sentences, due to the compositional interaction with the *if*-clause. In the MT counter-example (13), the instance of A in the major premise again functions to restrict the *ALL*-quantifier over situations provided by *always*. This is a very different job, semantically, from the use of A in the conclusion (13c) to predicate something of the current speech situation. A careful semantic analysis of these arguments reveals clearly why they are not intuitively compelling: both trade in a fallacy of equivocation. (As an exercise, the reader may wish to construct a similar analysis of the CD counter-example in (9).)

We are left with several possible interpretations of the situation so far. One option is to stick to our guns. The counter-examples were derived by uniform substitution of variables in the schematic argument forms by grammatical strings of English. Since the result of such substitution sometimes results in clearly

[2] The notation $Q[R][N]$ is borrowed from generalized quantifier theory [1, 31]. It is interpreted as follows: a quantifier Q takes two arguments, a restriction R and a nuclear scope N. For example, *All white dogs bark* would be translated as $ALL[\mathbf{white} \cap \mathbf{dog}][\mathbf{bark}]$, a claim that is true just in case the things that are both white and dogs are a subset of the things that bark.

specious arguments, we have shown that these argument forms are invalid. Call this a **raw string-based** picture of validity. If this position is right, then MP, MT, and CD are all invalid. (Worryingly, so are even more basic principles like Reflexivity, A implies A: see below.)

At the opposite extreme is a semantic or **interpretation-based** picture of validity, where the objects that we reason over are full semantic interpretations, with no information about strings. The basic idea is that validity is a relation among model-theoretic interpretations, involving truth-preservation; on this picture, formal relations among strings are neither necessary nor sufficient for argument validity.

Even if we can stomach the rejection of MP, MT, and CD, the raw string-based picture has some glaring problems. We did not quite use uniform substitution above, but rather replaced names by pronouns in some cases, assuming the reader would be able to fill in the appropriate interpretation (!) of the pronoun in each case. And we assumed that each token of ambiguous or polysemous words would be resolved in the same way. In addition, we relied on general principles of context-sensitive interpretation to determine, for example, when and where it is said to be raining/sunny/etc. [22]. When the strings we substitute have context-sensitive interpretations, even the simplest instances of MP, MT, and CD would be invalid if the context were not resolved in a uniform manner across sentences. But there is nothing in the strings themselves that tells us when we should hold context fixed across sentences, and when we should not.

Still, there seems to be a sense among theorists that ambiguity, anaphora resolution, and temporal/situational indexicality are fairly superficial aspects of interpretation. Perhaps the string-based picture can be maintained once these issues are patched up somehow—say, using the notion of "logical form" that I'll describe below.

Several recent papers seem to assume something like this in arguing against MT and CD. The counter-examples are strikingly similar to the ones given above, but they rely on a generalization of Lewis' observation by Kratzer [16, 17]. Kratzer pointed out that *if*-clauses also restrict the domains of epistemic and deontic modal operators.

(15) If he is rude, Mike ought to apologize.

   a. **Wrong**: If he is in fact rude in the actual situation, the best situations (in general) are ones in which Mike apologizes.

   b. **Right**: *Among situations in which Mike is rude*, the best are those in which he apologizes.

The incorrect interpretation in (15a) would imply that—if Mike is in fact rude—he ought to apologize, even in possible situations that are better than actuality–ones in which he has not been rude! The correct interpretation (15b) allows for the plausible option that Mike ought not to apologize in situations in which he hasn't done anything wrong.

Epistemic operators behave similarly: for instance, the intuitive interpretation of (16) makes reference to a conditional probability measure restricted by the antecedent *the die is above 3*, rather than an unconditional probability.

(16) If the die is above 3, it's probably even.
  a. **Wrong**: If above 3, $P(\mathbf{even}) > .5$.
  b. **Right**: $P(\mathbf{even} \mid \mathbf{above\ 3}) > .5$.

(16a) would have it that if the die is *in fact* above 3, then our *current* information implies that it's probably even—even if we have no idea what the outcome was. Arguably this is a *possible* interpretation, but if so it is certainly much less prominent than the one in (16b).

Yalcin [32] uses the interaction of epistemic operators and *if*-clauses to argue that MT is invalid, pointing to specious arguments along the lines of (17).

(17) a. If the die is above 3, it's probably even.
  b. It's not the case that the die is probably even.
  c. Therefore, the die is not above 3.

If a fair die has been rolled and we know nothing about how it came up, (17a) and (17b) are both true; and yet we clearly cannot conclude (17c). The semantic diagnosis is the same as it was in our AQ-based arguments. The instance of *probably* in (17a) picks out a certain condition on a probability measure whose domain has been restricted by the antecedent—that is, the conditional probability measure $P(\cdot \mid \mathbf{above\ 3})$. In contrast, the instance of *probably* in (17b) picks out a condition on the corresponding unconditional probability measure, with an unrestricted domain.

Along similar lines, Kolodny & MacFarlane [15] give a counter-example to CD involving the deontic modal *ought*. Without going into the full detail of their now-famous "Miners' Puzzle", the argument has a schematic form that should look familiar from (9):

(18) a. If A, ought C.
  b. If B, ought D.
  c. Either A or B.
  d. Therefore, either ought C or ought D.

Kolodny & MacFarlane observe that this argument fails intuitively in a scenario that they describe, and conclude that both CD and MP are invalid. Here again, the puzzle is generated by the fact that the instances of *ought* in (18d) talk about what is best globally, while those in (18a) and (18) talk about what is best in different restricted domains.

In both cases, the problematic examples involve a compositional semantic interaction *between* the two clauses that have been substituted in the antecedent and consequent of a conditional. A compositional interaction *within* one of these clauses would not have created a problem, since the interaction would be maintained in each instance of a string when we substitute it into the variables of an argument template.

3 Logical Form to the rescue?

The examples adduced by Yalcin [32] and Kolodny & MacFarlane [15] would certainly undermine MT and CD if we thought of these principles as applying to arguments derived by uniform substitution of grammatical strings of English for sentence variables. But no one advocates a theory like this: it would be hopeless, rendering even the most basic principles of reasoning obviously invalid. For instance, Reflexivity—"A implies A"—fails on raw strings because of the potential for context shifts.

(19) a. He is thoughtful. (pointing to Barack Obama)
b. Therefore, he is thoughtful. (pointing to Jair Bolsonaro)

The problem here is about what counts as "uniform substitution of sentences" in an argument template like Reflexivity or MP. Using string identity as the criterion, MP would of course be invalid—and every other logical principle you can think of.

(20) a. If he is thoughtful, I'll vote for him. (pointing to Jair Bolsonaro)
b. He is thoughtful. (pointing to Barack Obama)
c. Therefore, I'll vote for him. (pointing to Jair Bolsonaro)

While reasoning over raw strings of English is a non-starter, the opposite extreme—reasoning over full semantic interpretations—is not much better for our purposes. According to the classic semantic definition of validity, an argument with premises $\{\gamma_1, \gamma_2, ...\}$ and conclusion δ is valid if and only if, for all situations s: if each of $\gamma_1, \gamma_2, ...$ is true in s, δ is true in s as well. The semantic perspective is attractive because it automatically enforces some basic requirements for assessing validity of arguments. Since it makes crucial reference to *truth*, anaphora, ellipsis, indexicals, ambiguity, polysemy, quantifier scope, etc. must be resolved before the definition can even be applied. Sentences cannot be assessed for truth-in-a-situation without first performing these tasks.

With that said, the purely semantic approach to validity is also unsatisfying for many logical, psychological, and computational purposes. The basic question that we started with—"What is the status of principles like MP, MT, and CD?"—can't even be formulated within a strictly semantic approach. These are questions about the inferential relations among sentences with a particular *form*. The semantic picture of validity can see what sort of semantic object each premise denotes (e.g., a set of situations or "possible worlds" making the sentence true). It does not have access to information about whether a particular premise was formulated as a conditional, and this information is needed before we can pose the question of whether an example counts as an instance of MP. In other words, a theory at the pure semantic extreme would be theoretically handicapped because it would lack the resources to distinguish genuine instances of MP like (3) from specious instances like (20). In order to do this, a theory needs to pay attention to relevant aspects of the syntactic form of sentences (but not too much) in addition to their interpretations.

Note that this criticism applies to purely semantic theories of inferencing like the one that Johnson-Laird [14] seems to be arguing for, *even if* the theory is correct about the basic psychological mechanisms of reasoning. Simultaneous attention to the syntactic form of an argument and its semantics is needed in order to determine which apparent instances of an argument are genuinely relevant to our theoretical questions, and which are pseudo-arguments.[3]

To make their arguments against MT and CD work, then, Yalcin and Kolodny & MacFarlane must be assuming that validity is a property of syntactically structured representations. But these must be representations that crucially import *some* aspects of the semantic interpretation, to get the right criteria of identity for uniform substitution of sentences in argument templates. In logic and linguistics this task is often pursued by annotating pronouns and lexically ambiguous expressions with unpronounced numerical indices. Crucially, these elements are part of the syntactic form, but the way they are distributed has a purely semantic motivation—we assign indices depending on whether the pronouns differ in reference, or two instances of an ambiguous word have the same meaning, when the sentence is semantically interpreted. Helping ourselves to this additional information allows us to separate the specious argument in (19) from unproblematic instances of Reflexivity.

(21) a. He_1 is thoughtful.
b. **Good conclusion:** Therefore, he_1 is thoughtful.
c. **Bad conclusion:** Therefore, he_2 is thoughtful.

Linguists and philosophers call these semantically enriched syntactic representations "logical form". On some theories, each sentence (as tokened in some context) has a grammatically privileged logical form. Logical forms vary from theory to theory, but they are usually assumed to contain unpronounced, semantically relevant information about how to resolve anaphors and ambiguities, fill in ellipsis sites, and disambiguate quantifier scope alternations, among other tasks. All of this information—and probably quite a lot more—is needed if we hope to formulate a reasoning theory that operates over syntactically structured representations, while avoiding classifying arguments like (19)-(20) as valid.

Here's the punch line. In light of the purported counter-examples described above, the question of whether MP, MT, and CD are valid depends essentially on a theoretical choice point: whether our logical forms include information about the quantificational domains of modals and AQs. If they do, then the problematic examples given in §1 are no longer problematic. The enriched representations with domain annotations would look something like this.

(22) a. If it's raining, Mike $\text{always}^{(R\cap\textbf{rain})}$ carries an umbrella.

[3] With that said, the theory proposed by Johnson-Laird [13, 14] is not actually a representative of the purely semantic extreme that the author appears to advocate in the surrounding prose. The theory's "models" contain syntactic annotations such as negation and ellipses, making them rather close in spirit to the "logical forms" that are widely employed in linguistic and philosophical work.

b. It's raining.

c. Therefore, Mike always^R carries an umbrella.

This is no more problematic than the Obama/Bolsonaro argument in (19): the temptation to treat it as an instance of MP stems from a failure to attend to the silent domain index. Once we do, we see that this argument is not derived from the MP template by uniform substitution by logical forms of the appropriate type. The same diagnosis accounts for the AQ- and modal-based counter-examples to MT and CD. On the other hand, if we assume that logical forms do not contain information about quantificational domains, then all of these arguments are legitimate counter-examples, derived from the relevant argument templates by uniform substitution.

The fate of our bedrock reasoning principles hangs in the balance. How can we decide? If there is a linguistically and/or psychologically privileged level of logical form, the issue is an empirical one. We just need to do the empirical and theoretical work to find out whether domain variables do indeed exist at the level of logical form. A template for this kind of work is provided by Stanley and Szabó [29, 30], who argue that the domain arguments of nominal quantifiers like *every* are explicitly present at logical form. These arguments likely extend to AQs, with the implication that all of the counter-examples we saw above are spurious if Stanley and Szabó are correct.

Many theorists have reasons to be skeptical about the linguistic or psychological reality of a privileged, syntactic level of logical form. If this skepticism is warranted, then the choice of whether or not to include domain variables as in (22) comes down to a modeling choice, made for theoretical or practical convenience. As an analogy, consider an important strand of computational work geared toward the Recognizing Textual Entailment task. One of the driving considerations of this literature is that reasoning over full semantic interpretations is computationally prohibitive, while string-based reasoning is relatively cheap (e.g., [25]). However, raw strings are not especially useful for this purpose for a variety of reasons, including those considered above. So, reseachers in this tradition have suggested a variety of ways to implement reasoning systems that work with something close to the syntactic form of sentences of natural language, while adding annotations indicating semantic properties that are relevant to the task at hand.

One influential proposal along these lines, stemming from logic and later adopted into computational work, is the "monotonicity calculus" [3, 12, 24, 25]. The basic idea is that we annotate strings of a natural language (say, English) with information about syntactic constituency and the monotonicity properties of the phrases. Here, the "monotonicity" of an environment means, essentially, whether it licenses valid inferences from a category to a subordinate category, a superordinate, or neither. For example, from the premise *A car went by* we can validly infer *A vehicle went by*, but not *A Honda went by*. So, the position of *car* is annotated with a $+$, indicating that it allows substitution by superordinates, but not subordinates. Phrases marked with $-$ do the opposite, allowing inference to subordinates but not superordinates. So, *No car went by* implies *No Honda went by*, but does not imply *No vehicle went by.*

(23) a. (A $(\text{car})^+$ $(\text{went by})^+)^+$

b. (No $(\text{car})^-$ $(\text{went by})^-)^+$

c. (Every $(\text{boy})^-$ $(\text{laughed})^+)^+$

Negation flips $+/-$ annotations, so that the opposite pattern of substitutions yields valid inferences:

(24) a. (It's not true that (a $(\text{car})^-$ $(\text{went by})^-$ $)^-$ $)^+$

b. **Good**: Therefore, it's not true that a Honda went by.

c. **Bad**: Therefore, it's not true that a vehicle went by.

(25) a. (It's not true that (no $(\text{car})^+$ $(\text{went by})^+)^-$ $)^+$

b. **Good**: Therefore, it's not true that no vehicle went by.

c. **Bad**: Therefore, it's not true that no Honda went by.

This approach has been very useful in computational tasks involving textual entailment. The necessary annotations are fairly straightforward to generate, and they allow us to construct large numbers of candidate valid inferences—and rule out many invalid inferences—without engaging in the laborious task of constructing a full semantic interpretation [24, 25].

Crucially, the monotonicity profiles of quantifiers are determined by their semantic properties, not their syntax. But importing this semantic information into an enriched level of syntax as in (23)-(25) is practically useful for certain purposes. Similarly, we might construe the choice of whether to include domain variables at logical form as a theoretical or practical convenience. If so, a disquieting conclusion would seem to follow: there is no fact of the matter about whether MP, MT, and CD are "really" valid. Validity *of an argument form* is a concept that only makes sense relative to a choice of logical form. If our theoretical proclivities make us inclined to want these principles to come out valid—or, if we have some other motivation for annotating logical forms with domain variables—we can adopt a picture of logical form that does the task. On the other hand, if we choose logical forms that do not represent domain variables, then MP, MT, and CD are invalid, in line with the arguments of Yalcin and Kolodny/MacFarlane.

The reader can decide which way to go here: embrace a privileged level of logical form, or relativize the question of whether certain key reasoning principles are valid to an unforced theoretical choice. Either way, the startling conclusion is that the status of certain very basic reasoning principles—even modus ponens!—depends on fairly abstruse theoretical issues in natural language syntax and semantics.

4 Generics: Silent but dangerous

Generic and habitual language provides a second advertisement for the importance of attending to compositional semantics. Certain sentences of English are ambiguous between a one-off interpretation and an interpretation that expresses a generalization about how certain kinds of events tend to go. For instance, *Bill takes a walk after lunch* can be understood in either way, depending on context.

(26) a. (Why isn't Bill here now?) Bill takes a walk after lunch.
b. (What happens in the play's 4th act?) Bill takes a walk after lunch.

(In most languages, simple present sentences can also be interpreted as describing an ongoing action. English, unusually, requires the present progressive for this: *What is Bill doing? —He is taking a walk.*)

In formal semantics, the difference between generic and non-generic readings is usually traced to a silent adverb of quantification *GEN* which is present at logical form [18].

(27) Birds have wings $\rightsquigarrow$ *GEN* [x is a bird] [x has wings]
"Generally, if something is a bird it has wings"

I use the $\rightsquigarrow$ symbol to indicate a semantic interpretation (a fairly rough one, chosen for expository purposes).

Like overt AQs, the domain argument of *GEN* is often contextually provided. (Here, s is a variable over situations.)

(28) I cycle to work $\rightsquigarrow$ *GEN* [I go to work in s] [I cycle to work in s]
"Generally, when it's time to go to work, I cycle"

In this example, *GEN* doesn't quantify over all situations: if it did, the sentence would imply that I spend most of my time cycling to work. The intuitive interpretation involves a restriction to the (thankfully rare) moments when I need to commute and have a choice of methods.

If *GEN* is a covert AQ, we would expect that it interacts in the compositional semantics with *if*-clauses in the same way that AQs do—and it does. Notice that the content of the *if*-clauses in these examples ends up in the restriction of the *GEN*-quantifier [7].

(29) a. Bartenders are happy if they get big tips.
GEN [bartender x, x gets a big tip in s] [x is happy in s]
b. If it's raining, I take the bus to work.
GEN [it's time to go to work in s, it's raining in s] [I take the bus in s]

There has been a lot of work in recent years about the social and cognitive dangers of generic language (e.g., [4, 9, 21]). To these admittedly more pressing issues, we can add several further dangers specifically for researchers interested in reasoning. Because of their linguistic properties generics can mislead us in at least three ways:

1. *GEN* is invisible but has major semantic effects. We have to examine each sentence carefully in order to discern whether it is present.
2. *GEN*'s contextual restriction is not fully predictable from the linguistic form.
3. *GEN* interacts compositionally with other expressions, including *if*-clauses.

Dangers 1 and 3 are particularly relevant to research on conditional reasoning. Any given instance of a conditional might, if it is generic, involve complex semantic interactions that are absent in overly similar sentences that are not generic. To illustrate why this is a problem, consider Douven's [5] arguments against the three-valued semantics for conditionals proposed by de Finetti [8]. In this theory, a conditional is true if it has a true antecedent and true consequent; false if it has a true antecedent and false consequent; and otherwise undefined. Douven points out that this semantics has the consequence that certain right- and left-nested conditionals should have the same interpretation as simple conditionals with a conjunctive antecedent.

(30) a. **Right-nested**: If A, then (if B then C)
b. **Left-nested**: If (B if A), then C
c. **Conjunctive**: If (A and B), then C

The equivalence between (30a) and (30c)—the so-called "Import-Export" property—is widely thought to be an empirically correct prediction. However, Douven points to a number of examples with in which left-nested conditionals are clearly not identical to the apparently matched conjunctive conditionals. For instance,

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

Similarly, the three-valued de Finetti semantics predicts that a left-and-right-nested conditional of the form in (32a) should be equivalent to a three-conjunct simple conditional with the form in (32b).

(32) a. If (B if A), then (if C then D)
b. If (A and B and C) then D

Douven gives the following counter-example, where the two sentences clearly have different truth-conditions.

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

In light of these and similar examples, Douven concludes that de Finetti's theory is "materially inadequate because it gets the truth conditions and probabilities of nested conditionals badly wrong". However, Lassiter & Baratgin [19] point out that all of Douven's counter-examples involve generic sentences. So, a possible way out for an advocate of the de Finetti theory is to claim that non-generic sentences behave as predicted, but the generic examples differ for principled semantic reasons. Specifically, on their most natural interpretations the sentences in (31) have two key differences: how many instances of *GEN* there are, and where; and how the *if*-clause(s) interact with *GEN*.

(34) a. (31a) $\rightsquigarrow$ GEN (if (GEN B if A) then C)
b. (31b) $\rightsquigarrow$ GEN (if (A and B) then C)

If these are the logical forms that we construct in interpreting these sentences, they are not substitution instances of the templates in (30b) and (30c). As a result, de Finetti's theory does not predict that they should be equivalent. Indeed, when de Finetti's theory is combined with off-the-shelf theories of genericity [18] and domain restriction [11], the predicted truth-conditions are sensible renditions of the intuitive interpretations of these sentences in (32) and (33) (see [19] for details).

Lassiter & Baratgin note in addition that de Finetti's prediction of an equivalence between (30b) and (30c) is rather more plausible if we use sentences that are incompatible with a generic interpretation.

(35) a. If this material became soft at 3:05PM if it got hot at 3:04PM, our workers were not able to use it at 3:10PM.
b. If this material got hot at 3:04PM and became soft at 3:05PM, our workers were not able to use it at 3:10PM.

The use of specific time adverbials forces an episodic, non-generic interpretation, and the logical forms of these sentences are genuine instances of the templates in (30). Similarly, de Finetti's prediction that episodic left-and-right-nested conditionals are equivalent to thrice-conjunctive conditionals turns out to be very plausible when we enforce an episodic interpretation, ensuring that the example is a genuine substitution instance of the template in (32).

(36) a. If your mother got angry if you came home with a B last Thursday, then she'll get furious if you come home with a C tomorrow.
b. If you came home with a B last Thursday and your mother got angry and you come home with a C tomorrow, then your mother will get furious.

Attention to the non-obvious logical form of generic sentences, and to the compositional interaction of generic interpretation and *if*-clauses, is crucial to recognizing the difference between a spurious refutation like (31) and the genuine instance of the pattern like (35). The key remaining question is whether the sentences in (35) and (36) are indeed equivalent. In unpublished work with Maxime Bourlier, Baptiste Jacquet, and Jean Baratgin [20], we address this question in two experiments by asking participants to read pairs of sentences like (31) and (35). In the first experiment, participants made a binary judgment about whether the meanings were the same or different. In the second experiment, a different set of participants rated the same pairs on a continuous slider from

"Absolutely the same" meaning to "Absolutely different". Each pair of sentences was either designed to receive a generic interpretation, or decorated with temporal adverbials in order to force a generic interpretation. We also manipulated whether the first sentence was left-nested, right-nested, or left-and-right nested.

Combining de Finetti's semantics with linguistic theories of genericity and domain restriction, these are the key predictions:

1. Left-nested conditionals like (30b), when episodic, should be rated as better paraphrases of their purported conjunctive equivalents like (30c) than the matched generic pairs.
2. Left-and-right-nested conditionals, when episodic, should be rated as better paraphrases of their purported conjunctive equivalents than the matched generic pairs.
3. The generic/episodic manipulation should make no difference when right-nested conditionals are compared to their conjunctive paraphrases.

Both experiments bore out the three predictions. When the first sentence was right-nested, participants largely rated the pairs as having the same meaning regardless of genericity. In contrast, participants gave significantly more "same" judgments for episodic right-nested conditionals than for their generic counterparts in experiment 1, and ratings significantly higher on the sameness scale in experiment 2. The same patterns held for episodic vs. generic left-and-right nested conditionals.

The data are consistent with the predictions described above. Far from refuting the de Finetti theory, the interpretation of nested conditionals seems to be precisely what that theory would predict when combined with insights from modern linguistics. But the road to this conclusion was fraught with danger. Due to *GEN*'s invisibility, and because of its compositional interaction with *if*-clauses, we would have ended up with an erroneous picture of the logic of conditionals if we had not paid close attention to the insights of linguistic research. In the voluminous logical and psychological literature on conditional reasoning, it remains to be seen where else incorrect conclusions may have been reached due to a failure to attend to the complexities of generic interpretation. Psychologists beware!

5 Conclusion

Linguistic semantics and psychology of reasoning have a lot to learn from each other. Focusing on conditionals, this paper has discussed two examples where careful attention to the details of compositional semantics is crucial in making sense of confusing patterns, and in separating genuine problems from pseudo-problems. Of course, there are many areas for fruitful interaction beyond the study of conditionals (e.g., [10, 26]). While I have focused on ways that reasoning can benefit from the insights of linguists, the flow of insights in the opposite direction—both methodological and theoretical—is no less crucial.

References

1. Barwise, J., Cooper, R.: Generalized quantifiers and natural language. Linguistics and philosophy **4**(2), 159–219 (1981)
2. Bennett, J.F.: A Philosophical Guide to Conditionals. Oxford University Press (2003)
3. van Benthem, J.: Essays in logical semantics. Springer (1986)
4. Berio, L., Musholt, K.: How language shapes our minds: On the relationship between generics, stereotypes and social norms. Mind & Language (2022)
5. Douven, I.: On de Finetti on iterated conditionals. In: Computational models of rationality: Essays Dedicated to Gabriele Kern-Isberner on the occasion of her 60th birthday. pp. 265–279. College Publications (2016)
6. Evans, J.S.B.T., Over, D.E.: If. Oxford University Press (2004)
7. Farkas, D.F., Sugioka, Y.: Restrictive if/when clauses. Linguistics and Philosophy pp. 225–258 (1983)
8. de Finetti, B.: La logique de la probabilité. In: Actes du congrès international de philosophie scientifique. vol. 4, pp. 1–9. Hermann Editeurs Paris (1936)
9. Gelman, S.A.: Generics in society. Language in Society **50**(4), 517–532 (2021)
10. Geurts, B.: Reasoning with quantifiers. Cognition **86**(3), 223–251 (2003)
11. Huitink, J.: Modals, Conditionals and Compositionality. Ph.D. thesis, Radboud University Nijmegen (2008)
12. Icard III, T.F., Moss, L.S.: Recent progress on monotonicity. Linguistic Issues in Language Technology **9**, 167–194 (2014)
13. Johnson-Laird, P.N.: Mental Models. Cambridge University Press (1983)
14. Johnson-Laird, P.N.: Against logical form. Psychologica Belgica **50**(3), 193–221 (2010)
15. Kolodny, N., MacFarlane, J.: Ifs and oughts. Journal of Philosophy **107**(3), 115–143 (2010)
16. Kratzer, A.: Conditionals. In: von Stechow, A., Wunderlich, D. (eds.) Semantik: Ein internationales Handbuch der zeitgenössischen Forschung, pp. 651–656. Walter de Gruyter (1991)
17. Kratzer, A.: Modality. In: von Stechow, A., Wunderlich, D. (eds.) Semantik: Ein internationales Handbuch der zeitgenössischen Forschung, pp. 639–650. Walter de Gruyter (1991)
18. Krifka, M., Pelletier, F., Carlson, G., ter Meulen, A., Chierchia, G., Link, G.: Genericity: An introduction pp. 1–124 (1995)
19. Lassiter, D., Baratgin, J.: Nested conditionals and genericity in the de Finetti semantics. Thought: A Journal of Philosophy **10**(1), 42–52 (2021)
20. Lassiter, D., Bourlier, M., Jacquet, B., Baratgin, J.: How generics obscure the logic of conditionals (2023), ms., University of Edinburgh and University of Paris
21. Leslie, S.J.: The original sin of cognition. The Journal of Philosophy **114**(8), 395–421 (2017)
22. Lewis, D.: Index, context and content. In: Philosophical Papers, volume 1: Papers in Philosophical Logic. Cambridge University Press (1997)
23. Lewis, D.: Adverbs of quantification. In: Keenan, E.L. (ed.) Formal Semantics of Natural Language, pp. 178–188. Cambridge University Press (1975)
24. MacCartney, B.: Natural language inference. Ph.D. thesis, Stanford University (2009)
25. MacCartney, B., Manning, C.D.: Natural logic for textual inference. In: Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. pp. 193–200 (2007)

26. Mascarenhas, S., Koralus, P.: Illusory inferences with quantifiers. Thinking & Reasoning **23**(1), 33–48 (2017)
27. McGee, V.: A counterexample to modus ponens. The Journal of Philosophy **82**(9), 462–471 (1985)
28. Sinnott-Armstrong, W., Moor, J., Fogelin, R.: A defense of modus ponens. The Journal of Philosophy **83**(5), 296–300 (1986)
29. Stanley, J.: Context and logical form. Linguistics and philosophy **23**(4), 391–434 (2000)
30. Stanley, J., Gendler Szabó, Z.: On quantifier domain restriction. Mind & Language **15**(2-3), 219–261 (2000)
31. Szabolcsi, A.: Quantification. Cambridge University Press (2010)
32. Yalcin, S.: A counterexample to modus tollens. Journal of Philosophical Logic **41**(6), 1001–1024 (2012)