If you can’t get rid of the skeleton in your closet, you’d best teach it to dance.

George Bernard Shaw

With the benefit of hindsight, it is a bit surprising that “some” and “or” are the drosophilae of modern experimental pragmatics because the conditional (if-then) was primed to be the most studied logical expression. Researchers in the psychology of reasoning – even as far back as the 1970s and 1980s – were focused on the way children and adults represent and reason with conditionals. These efforts raised questions about the role associated with pragmatic enrichments. In the meanwhile, theorists in the linguistic-pragmatic literature, as we just saw in the last chapter, were taken by the way pragmatic enrichments arise when embedded in the antecedent of conditionals. Theorists also considered how the connective “if” often appears to mean the more informative “If and only if” (see Chapter 5). Interestingly, the reasoning and pragmatics literatures hardly intersected with respect to conditionals (e.g., the developmental data that enlisted the notion of invited inferences in Chapter 5 from Rumain et al. [1983] represent an exception rather than the rule).

What did each of these two literatures report? The psychological literature largely treated conditionals as a set of two valid and two invalid inferences (as described in Chapter 5). As a reminder, the two valid inferences are Modus Ponens (If P then Q: P//Therefore, Q), as represented in (8.1), which is generally carried out routinely (children and adults evaluate the conclusion, below the line, as true at rates of at least 90 percent) and Modus Tollens (If P then Q: not-Q//Therefore, not-P), which is shown in (8.2), which is not carried out with ease and leads to middling rates of correct responses.

(8.1) If the fish is red then it is striped.
The fish is red.
The fish is striped.

(8.2) If the fish is red then it is striped.
The fish is not striped.
The fish is not red.
Modus Tollens is difficult because in order to evaluate the conclusion one needs to first make the supposition that the antecedent \( P \) (the fish is red) is true and notice that this would lead to a conclusion (the fish is striped) that is in contradiction with the minor premise. This amounts to a reductio ad absurdum argument that prompts the rejection of the supposition \( P \), the conclusion in (8.2).

The two invalid (fallacious) inferences amount to two pragmatic inferences that will concern us in this chapter. These are Affirmation of the Consequent in (8.3) and Denial of the Antecedent in (8.4), which were described in Chapter 5 when I mentioned Geis and Zwicky (1971).

(8.3) If the fish is red then it is striped.

\[
\text{The fish is striped.} \\
\text{The fish is red.}
\]

(8.4) If the fish is red then it is striped.

\[
\text{The fish is not red.} \\
\text{The fish is not striped.}
\]

The conclusion in the Affirmation of the Consequent (AC) argument is accepted as true by anywhere from 20 percent to 80 percent of participants and across a range of conditional reasoning tasks (for slightly out of date summaries, see Evans, 1993, or Girotto et al., 1997, but newer data are within this range too). One finds similar rates with respect to the conclusion in the Denial-of-the-Antecedent (DA) syllogism. From a logical point of view, the minor premises in (8.3) and (8.4) are non-starters for prompting further inferences. That is, once a participant encounters one of these sorts of minor premises (in the context of a conditional), the argument is invalid and the logical response is something along the lines of “nothing follows.” However, as the data indicate, that is not how participants see things. Participants often carry out the invalid inference and are, at best, mixed about responding logically.

As we saw in Chapter 5, Geis and Zwicky (1971) suggested a principle known as Conditional Perfection, which would explain how a conditional (if-then sentences) could be seemingly interpreted as a biconditional (If and only if, see Chapter 5, fn 8), which justifies the pragmatic inferences in (8.3) and (8.4). Over the years, pragmatists aimed to better account for “perfection” and they would invariably start by employing a familiar approach – Gricean maxims and scales. The thinking was (and largely continues to be among linguists) that the conditional implicitly generates another proposition that can be considered more informative than the uttered conditional (and that can ultimately be rejected). The question has been, what are the more informative renderings of if-then?
One suggestion came from van der Auwera (1997), who, in a review that covers 25 years of discussion, argued in favor of his own (and what he called tongue-in-cheek the “correct”) view. As can be seen below, he begins by defending an ideal based on scales before presenting his suggestion:

In my view, the explanation of the conditional perfection phenomenon is Gricean. We are dealing with a scalar Quantity implicature. The type of scale involved is [the one represented in (8.5)]. The higher assertions entail the lower ones. Standard scalar implicatures arise as negations of the higher assertions and this is what we find here ... One thus implicates that only $p$ will do and that if $p$ is not fulfilled, $q$ will not ensue.

\[(8.5) \quad \ldots \]
\[
\begin{align*}
&\text{if } p, q \text{ and if } r, q \text{ and if } s, q \\
&\text{if } p, q \text{ and if } r, q \\
&\text{if } p, q
\end{align*}
\]

van der Auwera’s idea is that when one says, for example, “If the fish is red then it is striped” in (8.1), this is relatively underinformative; instead, according to van der Auwera, the speaker could have said, “If the fish is red then it is striped and if it is blue then it is striped” but did not. This more-informative assertion can then be rejected (it would become “if blue then not striped”), putting one in the position to draw the conclusion that the speaker’s only condition for stripedness is the color red. This turns the conditional into the biconditional “If and only if the fish is red then it is striped.” While many share van der Auwera’s approach, other theorists (including Horn, 1972, and Matsumoto, 1995) consider his choice of “higher assertions” to be too complex (too “prolix”), which would violate another of Grice’s maxims (compare this to the case of “some,” where the rejected more-informative-alternatives are other highly frequent words). Horn (2000) later suggests that a better higher assertion might simply be $q$ or, alternatively, as von Fintel (2001) put it, $q$ no matter what (see also Ducrot, 1972; Haspelmath & König, 1998).3 As can be seen, the distinction between van der Auwera’s account and the others concerns the nature of the scale, but the adherence to scales is obviously a strong one.

A relevant side note is that there was universal agreement about discounting the idea that one might be tempted to consider the biconditional as a more informative alternative. After all, it entails the conditional (much like All is a good entailing alternative for Some):

\[(8.6) \quad \text{if and only if } p \text{ then } q \]
\[
\begin{align*}
&\text{if } p, \text{ then } q
\end{align*}
\]

The problem with this possibility is that – if one assumes that the higher assertions are there to be negated – one would actually be negating the very
proposition that theorists are seeking to justify. The way theorists justify this non-move is to say, for example, that this scale is ill-formed (e.g., Atlas and Levinson say that this scale relies on a higher assertion that is complex) or that if-and-only-if is not a word, so it cannot serve as an alternative.

In light of these objections and complications, Horn (2000) eventually proposed that, unlike the scalar terms Some and Or, which rely on a Quantity- (Q-) based maxim, a conditional is strengthened to a biconditional through a different, Relevance (R-) based maxim. That is, the weak conditional gets narrowed to mean the biconditional, much like the way the general term drink is used more specifically to describe an alcoholic beverage or the way “I don’t believe that p” is meant to be understood more narrowly as “I believe that not-p.” In the case of conditionals, enrichments do not occur due to a trimming off of stronger terms, they are simply narrowed directly to a more refined meaning.4

Although the two classes of linguistic proposals are at odds about the way the interpretation comes about, theorists end up with some form of conditional perfection – i.e., the conditional is transformed into a biconditional. In its original form, the antecedent in the conditional is simply sufficient for the consequent; once the biconditional transformation is complete, the antecedent acts as a necessary and sufficient condition for the consequent. Note that theorists are trying to come up with an explanation that will account for, what strikes me as, an intuition. Again, this reliance on intuition is exactly what an experimentalist guards against and what I will address later. Before getting to relevant experiments, I first want to quickly address another question.

Why did the two (linguistic-pragmatic and psychology of reasoning) literatures fail to seriously engage with one another? My humble take on this question is that, while there were a number of people who individually recognized an interesting overlap across the two literatures (e.g., Leo Noordman and Martin Braine), there was not a critical mass of investigators asking the questions that would sustain interest among both fields. The reasons for this are twofold. On the psychological side, most specialists of reasoning (those who would be concerned with if-then) did not take linguistically oriented concerns seriously enough. Instead, they feuded over cognitive representations (“Are inferences carried out by logic-neutral models or through logical schemas?”) that did not speak to linguists. Moreover, psychologists became interested in non-linguistic factors, such as the way low-level strategies can undermine logical inference making.5 On the linguistics side, scholars did not sustain their interest in conditionals and when they did they rarely cited psychological data, as far as I can tell. The time was just not ripe, I suppose.6

If the two literatures had collaborated earlier, the way linguists and psychologists do now about scalars, conditionals could very well have overshadowed scalars as the main topic of experimental pragmatics. After all, the conditional has much to offer. The equivocality among adults in the psychological literature
on conditionals showed promise just like the early scalar data. Conditionals also prompt reliable developmental data. Conditionals could have engendered a linguistic debate between several opposing Gricean-inspired theoretical descriptions of conditional enrichments as well as discussions among those who rely on scales (and the quantity maxim) versus those who rely on R-based accounts.

Imagining an alternative history is fun to think about, but it is just that – a counterfactual. The upshot is that much research remains to be done on conditional processing and we have a large set of data from scalars that can now serve as background. In the next section I summarize the data on conditionals that went largely overlooked by pragmatists. Along the way, I turn to work carried out in my own lab since and partly to show how a concerted group of experimentalists can attack a relevant pragmatic issue. I think the experimental pragmatic literature on the semantic/pragmatic divide on conditionals would look very different today if other groups were to pursue this difficult case.

**Enrichments Linked to Conditionals Do Not Line Up With Those of Scalars**

Although much of the data collected on the comprehension of conditionals are not recent, they use familiar techniques (involving development, reading times, and individual differences) and varied paradigms. In some cases, participants are asked to evaluate conditional rules based on presented evidence and at other times participants are required to evaluate conclusions from syllogisms. I will review some of these studies to provide flavor and, where relevant, I will juxtapose results with those reported in the previous chapters about scalar terms. By the time we get to the end of the section, it should be clear that conditional perfection (enriching \( \text{if} \) to \( \text{if and only if} \)), when it occurs, is actually a different kind of pragmatic enrichment, despite accounts that liken them to scalars.

**The Development of the Comprehension of Conditionals**

We have seen that the semantic meaning of scalar terms is often acceptable for young children and that pragmatic enrichments become evident with age (while keeping in mind that tasks vary in their difficulty). If this phenomenon were applied to conditionals, one would expect younger children to apply a logical reading with relative ease (e.g., correct evaluations to Modus Ponens arguments) and to increasingly apply pragmatic readings – to accept AC and DA arguments – with age. This is hardly the story. In fact, advancing age and further effort are linked to fewer pragmatic interpretations with conditionals (Barrouillet et al., 2000; O’Brien et al., 1989; Taplin et al., 1974). These studies reveal that children are more likely to accept AC and DA inferences
when they are younger before becoming more circumspect with age by being more likely to reject the conclusions in arguments such as (8.3) and (8.4). Ultimately, adults end up in both columns by providing both logical and non-logical responses.

To give one example in detail, consider O’Brien et al. (1989) who showed how conditionals are understood among seven-year-olds, ten-year-olds, and adults. Imagine you are a participant who is shown four boxes, each containing one animal and one fruit, as schematically represented in Figure 8.1; there is one box with a toy cat and banana, one with a toy dog and an orange, one with a toy dog and an apple and, finally, one with a toy horse and an apple. Actually, participants were shown two sets of these four boxes. One set was covered and placed on the side momentarily while the other would remain uncovered in front of the participant as a reminder of what the covered boxes could look like. During the test phase, a single, randomly chosen covered box would be presented. This was the basis for (what could be) dozens of true or false conditional statements. About a single box, one can truthfully say that, “If there is a banana then there is a cat” as well as the inverse “If there is a cat then there is a banana.” Dissimilarly, one can truthfully say that, “If there is a horse then there is an apple,” but not the inverse. Developmental trends show that nearly all seven-year-olds, who answer control sentences correctly (e.g., by saying false to “If there is a horse in the box then there is a dog”), say true (incorrectly) to “If there is an apple in the box then there is a horse” and “If there is a dog in the box then there is an orange.” Adults answer these correctly and the ten-year-olds are intermediate. It is as if younger children treat “If there is an x in the box then there is a y” as “There is a box that has an x and a y,” which means that they do not readily apply a conditional relationship between the objects. Younger children understand “If” as a conjunction which mimics the acceptance of the AC arguments above.

These data also show that the youngest children do not treat the conditional as a biconditional. If they did, they would say false to “If there is an orange then there is a dog” because the inverse is false; they say true, because they notice that “an orange” is joined with “a dog” in a box and that is good enough evidence for saying true. This research on conditionals shows that, as kids get older, they better consider the range of evidence that the antecedent calls upon, which leads to improved performance.

![Figure 8.1. A schematic representation of four boxes containing toy animals and fruits that serves as basis of a conditional inference task](https://www.cambridge.org/core/core.png)
Reaction Time Data

As far as I know, no one in the conditional perfection literature has made a claim that says if-then is understood by default as a biconditional. On the other hand, it is plausible to suppose, based on our prior findings on scalars, that theorists would not object to a proposal that says conditional perfection is more effortful than a reading that does not call for it. Arguably, pragmatic enrichments of conditionals can be processed in a way similar to enrichments linked to scalars, where good-enough semantic readings of Some, for example, are typically processed more easily than those that are revealing of pragmatic enrichments.

This is the hypothesis we investigate here. To keep things simple, from here on I will focus on participants’ reactions to AC arguments as in (8.3) and Modus Ponens (MP) arguments, as in (8.1) because these two arguments have many of the same elements and they come with no negations (while the minor premise of MP is the conditional’s antecedent and its conclusion is the consequent; the minor premise of AC is the conditional’s consequent and its conclusion is its antecedent). The only difference between the two arguments is that saying true to MP does not require conditional perfection, while saying true to AC does. In other words, I will consider participants’ performance on MP as a control for performance with AC. The other option for a control would be the normative (logical) answer to AC – “can’t tell” or “insufficient information to decide” – but these are clearly harder to capture quickly when compared to a simple “yes” or “true.” That said, I will continue to note the rates of logically correct performance on the AC argument, which remains an important reference.

With this in mind, let’s consider two of the older reaction time studies that concern conditionals. First up is Marcus and Rips (1979, Experiment 2), who presented the four sorts of conditional syllogism – Modus Ponens, Affirmation of the Consequent, Denial of the Antecedent, and Modus Tollens – and recorded their participants’ reaction times. While all of their participants endorsed MP conclusions and only about a third of them endorsed AC arguments, their data showed that the speeds at which the two are endorsed are statistically comparable (1,813 ms to endorse MP and 1,907 ms, to AC). Importantly, rejections of AC arguments, coinciding with a logical interpretation of the conditional, were more common and took noticeably longer to carry out (2,437 ms).9

In contrast, Barrouillet et al. (2000), who investigated the endorsement patterns of all four conditional arguments, found rates of AC endorsements (79 percent) that are among the highest in the literature; more importantly, the associated response times (2,355 ms) were slightly, but significantly, slower than MPs (2,018 ms). Notably, their experimental design, in which participants were required to hold a conditional premise in memory as a series of minor premise-conclusion combinations were presented, was slightly different. Also,
reaction times of those who rejected conclusions from an AC argument were not reported (probably because there were not enough of them to produce statistically reliable means). So, while this experiment’s design is more complex than Marcus and Rips’s, it draws the conclusion that affirming the consequent can appear to be more effortful than Modus Ponens.

This means that the data by the turn of the century was mixed on the question of whether or not an invited inference is effort-demanding. Note that the early efforts in the scalar literature were unequivocal. As noted in Chapter 6, all the early attempts showed that enriched readings came with longer reading times. As will become clear in the next section, work conducted by my colleagues and myself aimed to, among other things, determine which of the two empirical claims is better supported.

**Individual Differences**

When individual differences have been reported with scalars, they concern distinctions between those who are more or less skilled pragmatic hearers (e.g., through Theory of Mind measures). Dissimilarly, individual differences linked to performance with conditionals concern other cognitive factors such as memory abilities and IQ (De Neys et al., 2005; Evans et al., 2007, 2008, 2009). For example, De Neys et al. (2005) gave participants a memory task (developed by La Pointe & Engle, 1990), in which participants were given a mathematical operation followed by a word (15 trials all told). The mathematical operation would be, for example, $(4/2) - 1 = 5?$ and the word would be BALL. Participant memory scores were based on how well they recalled the words and in their presented order. After removing participants who made too many errors on the math problems or who took too long to carry them out, De Neys et al. took participants from the highest and lowest quartiles and gave them conditional reasoning problems in the format exemplified by (8.7) before asking them how sure they are about drawing the conclusion by using a seven-point scale:

(8.7) Rule: If Ann turns on the air conditioner, then she feels cool.
Fact: Ann turns on the air conditioner.
Conclusion: Ann feels cool.

Remarkably, those with high-memory spans were generally more likely to give ratings reflecting greater normativity. That is, compared to those in the lowest quartile for memory, those with the strongest memories were more sure about (provided higher ratings for) properly logical inferences, Modus Ponens (MP) and Modus Tollens (MT), while being less sure about the pragmatically valid Affirmation of the Consequent (AC) and Denial of the Antecedent (DA). Once again, superior normative reasoning performance on a basic conditional task.
is linked with enhancing logical tendencies and suppressing non-logical (pragmatic) ones.

Overall, the data concerning pragmatic enrichments with respect to conditionals tell a different story compared to those linked to scalars and across three sections of literature: (1) whereas children lean toward semantic readings of scalar terms and become progressively “pragmatic,” children’s early interpretations of conditionals appear compatible with pragmatic interpretations but become progressively logical (at which point the pragmatic interpretations appear to be blocked, discarded, or unconsidered); (2) whereas processing times of scalar items among adults are linked with the presence of pragmatic enrichments, extra effort – when it makes itself known – appears linked to avoiding the so-called invited inferences that often appear with conditionals (in one of two studies, one notes a longer reaction time to an endorsed AC argument, when compared to MP, but in another study, longer reaction times are linked to the rejection of an AC argument, as opposed to its acceptance); and (3) whereas differences in scalar enrichment have inspired investigations into measures of Theory of Mind, enrichments linked to conditionals deal with memory abilities or intelligence scores. When viewed through the prism of pragmatic enrichment behavior described in Chapters 6 and 7, the pragmatic enrichment of conditionals seems characteristically different, even though many linguistic-pragmatists use the very same tools – the Quantity maxim and scales – to describe both phenomena. Something has to give.

Doing Experimental Pragmatics on Conditionals

With this paradox as background, I turn now to work on conditional processing that my colleagues (principally Mathilde Bonnefond and Jean-Baptiste Van der Henst) and I carried out in Lyon. It is work that has not received nearly as much attention as the work on scalar terms, but is based on the very same approach that helped make “some” and “or” the focus of so much attention. Plus, we conducted many of these studies with the benefit of hindsight about the processing of scalar terms with the express purpose of determining whether the lessons drawn from scalars can be generalized to the case of conditionals. One of our main goals was to resolve the contradictory findings concerning reaction times (as a reminder, we decided to focus uniquely on the AC argument because it is negation-free). Assuming that AC confirmations are pragmatic enrichments, should they not prompt extra effort compared to controls (in this case, Modus Ponens)? Before I get to the findings, it is useful to describe our approach, which relied on perseverance and great teamwork.

Our lab in Lyon wanted to come up with a paradigm that would put us in a position to accomplish three things. First, we intended to set up an EEG experiment on conditional processing, which entails a word-by-word paradigm. So,
we needed to come up with a way to eventually isolate, not only each line, but each word in a syllogism and we wanted to be sure that such a presentation would not be detrimental to our data collection. When we began, our focus was on the last word of each syllogism, because we had figured that a participant makes a final judgment then. Second, we intended to use reaction times, too, to make hypotheses about how participants evaluate the MP and AC arguments, and this is best done in more naturalistic settings (i.e., outside the EEG chamber without electrodes on the head and without a focus on individual words). This implies a more ecologically valid sentence-by-sentence (as opposed to word-by-word) presentation. Finally, we wanted to see whether participants differ in the way they answer and to determine how reading times pattern with these individual differences.

We also kept in mind the prior studies (described earlier), which measured reaction times from either entire syllogisms (Marcus & Rips, 1979) or from the last two lines, the minor premise and the conclusion, while the major premise was committed to memory (Barrouillet et al., 2000). We, obviously, were taking this progression several steps further by breaking down the arguments into single lines and eventually words. As part of this intensive undertaking, Jean-Baptiste came up with syllogisms that looked like the Modus Ponens argument (8.8) below:

(8.8) If Jean goes to the cinema then he travels by bicycle.
    Jean goes to the cinema.
    He travels by bicycle.

These sorts of slice-of-life syllogisms made it easy to come up with control conditions that could keep participants honest through the replacement of a single word; that is, we also presented control items with conclusions that can render a syllogism false, such as, “He travels by bus,” which would replace the conclusion in (8.8). Of course, it makes sense to provide the AC argument too.

(8.9) If Jean goes to the cinema then he travels by bicycle.
    Jean travels by bicycle.
    He goes to the cinema.

This type of argument could be presented with a thoroughly anomalous conclusion such as “He goes to the gym,” which is inconsistent with the invited inference. We called (8.8) and (8.9) Modus Ponens-Consistent (MP-Consistent) and Affirmation of the Consequent-Consistent (AC-Consistent), respectively. Through the replacement of the last word, we came up with conditions that were “Inconsistent” for each.

Our first efforts addressed our preoccupation with carrying out an EEG study. We presented the four kinds of conditional arguments to two groups, either sentence-by-sentence or word-by-word, as we wanted to make sure that
the presentation did not unduly affect participants’ evaluations. To give an idea of how we went about this, the sentence-by-sentence group saw a left-justified cross midway up a screen to indicate that the upcoming line was a premise and, likewise, an arrow in the same place in order to signal a conclusion. Participants then had to make a choice between two options on the keyboard: *logical* or *not logical*. The word-by-word condition was also presented in the middle of the screen, but in the center; as before, a syllogism began with a cross to signal a premise and an arrow for the conclusion. This technique is known as Rapid Serial Visual Presentation (RSVP). Aside from the presentation, the main difference between the two was that the sentence-by-sentence condition gave two full seconds per line (with a 200 millisecond pause after each sentence) and the word-by-word condition showed each word for 400 ms (followed by a 200 ms pause after each word). In light of the fact that the materials were provided in French, the first sentence always contained 10 words and both the minor premise and the conclusion had 4. This translated to a difference in terms of uptake. The sentence-by-sentence condition gave participants 2.2 s to read each line (6.6 s in total) while the word-by-word condition gave participants 6 s to read the first premise and 2.4 s read the minor premise and the conclusion each (10.8 s in total).

To our mild surprise, there was one major difference between the word-by-word condition and the sentence-by-sentence condition. While participants correctly evaluated the two Modus Ponens arguments at rates approaching ceiling and they correctly rejected the thoroughly anomalous AC-Inconsistent condition, more people rejected what we called the AC-consistent condition when it came in the more measured word-by-word condition.

We considered these results an indication that the extra uptake time was encouraging more logically correct evaluations. This was interesting in itself and led to a short paper (van der Henst et al., 2006) and, as far as paradigm development was concerned, we concluded that our word-by-word technique did not undermine the performance that is usually found. While rates of correct performance shifted, we still found that participants treated the AC argument in one of two ways and that participants generally stuck to one of those ways throughout. I will refer to those participants who consider the AC-Consistent conclusion as “logical” as *Endorsers* and those who consider the AC-Consistent conclusion as “non-logical” as *Rejectors*. We were convinced that the word-by-word presentation could provide us with a way to study conditional arguments “on-line” through EEG without compromising typical performance. We held off from reporting the reaction time results because we began to realize that the conclusion, our intended target, was not providing the most interesting results.

As part of a Masters project in Lyon in 2004, Kinga Bujakowska, who now works as a molecular geneticist at Harvard University, tested participants with
the above word-by-word paradigm while using EEG. Kinga’s work provided us with a second mild surprise because she found that participants’ reacted differently to the second word of the second premise (the 12th word overall in French) of the AC syllogisms when compared to the 12th word of the MP syllogisms. For example, the word “travels” in the minor premise in (8.9) prompted negative going activity on the frontal area of the scalp 200–300 ms after its onset with respect to the word “goes” in the minor premise in (8.8). Note how this is the first word that indicates the direction in which the argument goes. I will address what this negative-going activity indicates in greater detail below, but for now I’ll say that it shows that participants detected something unexpected about the minor premise in the AC syllogism. All in all, we succeeded in isolating the very moment that distinguishes MP and AC inferences with a straightforward task. With this astounding finding in hand, we were ready to refine the paradigm further.

That was when the baton was passed to Mathilde Bonnefond, a biologist who had given up working with animals due to the onset of a severe allergy and who had, and has, an enduring passion for EEG (and MEG) technology. For her PhD thesis (under Jean-Baptiste’s direction), Mathilde began by further simplifying the materials in the above paradigm in the interest of having an even cleaner signal. Instead of having conditional arguments that present slice of life scenarios (about Jean’s plans), participants now saw MP-Consistent and AC-Consistent problems as (8.10) and (8.11), respectively:

(8.10) If there is a B then there is a D.
   There is a B.
   There is a D.

(8.11) If there is a G then there is a K.
   There is a K.
   There is a G.

This way participants could ignore any extraneous information that can be construed from even trivially thematic premises and, as before, the Inconsistent problems could be created with a small change – the substitution of a letter. For example, a MP-Inconsistent problem could result from changing D in the conclusion of (8.10) to an E, and an AC-Inconsistent could be created by changing G in the conclusion of (8.11) into a T. Afterwards Bonnefond et al. (2012) ran two experiments.

One was a behavioral study that was carried out sentence-by-sentence. This time, however, the task was self-paced, meaning that participants moved the argument along with a press of the spacebar of an ordinary keyboard and answered “logical” or “not logical” through a single key (one labeled on the right side and another labeled on the left). As we had become accustomed, this
produced two sorts of participants. There were participants who consistently endorsed the AC-Consistent syllogism and others who did not (there were also “mixed” participants, but for the sake of simplicity, I will not address their data here). Remarkably, again, Endorsers process the minor premise in the AC syllogisms significantly more slowly than they do the minor premise in the MP arguments (see Figure 8.2).

Why is this finding relevant? Because none of the existing linguistic-pragmatic theories proposed can account for it (no matter what Gricean maxim is used as part of the explanation). If one assumes that the conditional gets enriched (so that it is interpreted as a biconditional), this should occur when the Major Premise is processed beforehand. Once enriched, it should not matter whether the conditional’s antecedent or consequent serves as a minor premise. Yet, it appears that it does matter, even for those participants who routinely endorse the AC argument. Also, one cannot assume that the enrichment of the conditional is akin to our other storied enrichment, viz. the scalar. Another finding from Bonnefond et al.’s first experiment (one that is not new) is that those who provide “logical” evaluations to the AC-Consistent syllogisms (Rejectors) take a longer time to get through the Minor Premise than those who endorse the AC-Consistent syllogisms (Endorsers). Thus, one cannot say that logical interpretations at this crucial juncture are more basic than the pragmatic one. No matter how one slices it, the Affirmation of the Consequent argument, which is the most ideal test of conditional perfection, appears to behave in a way that is unlike anything we have seen so far.

Capturing relative speeds of response allowed us to infer where extra processing is taking place. It did not allow us to determine the nature of the extra effort, which is what we turned to next. Bonnefond et al.’s Experiment 2 was practically identical to Experiment 1, except that it was designed for an EEG study and was thus modified slightly to save time as well as to come up with

<table>
<thead>
<tr>
<th>Major premise</th>
<th>Minor premise</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>If P then Q</td>
<td>P</td>
<td>Q (MP)</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>P (AC)</td>
</tr>
</tbody>
</table>

812 ms
854 ms

Figure 8.2. This summarizes the behavioral finding from Bonnefond et al. (2012) showing that even Endorsers, who fallaciously accept Affirmation of the Consequent (AC) arguments, take longer to process the minor premise of AC than they do the minor premise of Modus Ponens (MP).
a cleaner signal. Compared to the paradigm of Experiment 1, it (a) had more trials, (b) further reduced wordiness (i.e., “If there is an A then there is a C” became “If A then C” and “There is an A” became “A”) so that participants read while keeping eye movements to a minimum; and it (c) added filler items whose minor premise had nothing to do with the conditional so that participants would not come to expect either the minor premise for MP or AC (e.g., \( If \ P \ then \ Q; \ U \)). Our goal from the beginning was to develop a satisfying and thorough EEG experiment to investigate conditionals. We finally landed on our final paradigm after years of refining our methods.

The EEG study further supported what Kinga’s study had indicated earlier. Namely, the minor premise of an AC argument prompts what we would eventually determine to be an N200. What is an N200 (or an N2)? It is an ERP component that is related to the violation of expectations (for a review, see Folstein & Van Petten, 2008). To give an example of an N2 from another cognitive area, consider a working memory task that requires participants to determine whether a second stimulus (S2) is similar or not to an earlier stimulus (S1) presented a few milliseconds earlier; a mismatching S2 elicits a larger N2 component than a matching S2 component (Wang et al., 2002; Zhang, Wang, Li, & Wang, 2003). As in Experiment 1’s self-paced study, we found that the arrival of the minor premise for an AC argument prompts a universal reaction, an N2 that reveals a dispreference. Not only do we find it among participants who ultimately reject the AC argument, we find it among those who will later endorse it. That said, it is even stronger among Rejectors. This result is noteworthy because it is an indication that the minor premise of the AC argument is to some extent disruptive to all participants. Below (in Figure 8.3) is one of nine measures taken that captures the difference across conditions from the moment that the second premise of MP and AC is presented. The one on the left represents the midline frontal site among Endorsers, who ultimately accept the MP-Consistent and AC-Consistent arguments, and the one on the right captures the same site among Rejectors.

My view, as developed over the course of several generations of students and experiments, is that participants note that the minor premise of AC arguments is a source of conflict (Noveck et al., 2011). To wit, all participants (Endorsers and Rejectors) react similarly to (what is essentially) an unexpected minor premise; profiles differ only with respect to degree. There is thus little reason to suppose that the conditional premise lends itself — in any way, shape, or form — to conditional perfection, which presupposes a seamless enrichment. Once something goes awry (because one did not get a minor premise for MP at the moment it was expected), one can react in one of two ways. A generous reaction can ultimately lead to finding ways to endorse the argument (dashed expectations notwithstanding) while a strict reaction is to recognize the flaw and to not budge.
When one turns to scalars (see Hartshorne et al., 2015; Noveck & Posada, 2003, Nieuwland et al., 2010), one finds nothing like what we found here. Data from EEG studies with scalar inferences prompt very different reactions. Those who respond “logically” to “some” do not prompt steeper ERP components and when one finds something to report with respect to scalar enrichments, the component described is weaker, later, or requires relatively rich contextual information. In contrast, the logical responses with respect to AC (the Rejectors) prompt the steepest N200s (even when compared to Endorsers). This brought to a close a very intensive period of joint work. Our behavioral and ERP studies showed that conditionals prompt reactions that are categorically different from the intensively studied scalars. We could safely conclude that the notion of conditional perfection and its concomitant invited inferences are linguistic proposals that have no support. We are better off thinking about the pragmatic import of conditionals very differently, perhaps as a means to heighten expectations about the antecedent.

Notes

1. Modus Ponens can be “disabled” if an argument presents a second conditional premise. For example; Modus Ponens will not be carried out mindlessly in the following set of premises (Byrne, 1989):

   If she meets her friend then she will go to a play.
   If she has enough money then she will go to a play.
   She meets her friend.

Figure 8.3. Stimulus locked grand-average waveforms evoked by the appearance of the letter representing the Minor Premise of the MP (black line) and AC (gray line) conditions on the frontal midline site for Endorsers (N = 17) and Rejectors (N = 23). The N200s are shown with an arrow and indicate statistically significant effects (from Bonnefond et al., 2012)
2. While a biconditional interpretation accounts for both AC and DA, a conjunctive interpretation could account for AC. In fact, the conjunction entails the biconditional, which entails the conditional. This chapter will focus on AC but will continue to test linguists’ classic notion that the pragmatic enrichment of conditionals involve a “perfection” to a biconditional interpretation.

3. In my view, this would not encourage the acceptance of invalid arguments (and the production of perfection) but the opposite; it would encourage valid evaluations of “can’t tell.” Assuming that the conditional If P then Q prompts the rejection of (the higher assertion) Q, the minor premise Q (for example, in 8.3) would be cause for a contradiction and the immediate breakdown of the argument.

4. Note that this can apply to Some too; “only some” is a narrowing of Some.

5. Consider Matching bias, which refers to the non-normative performance that occurs when elements mentioned in a rule do not correspond with those in a test item (e.g., consider the double mismatch between the rule “If there is a not a T on the card then there is not a 4” and a card showing H6 (see Evans, 1998; Prado & Noveck, 2007).

6. For the record, investigations into conditionals remain vital to several lines of reasoning research. One concerns the way logically valid conditional inferences, such as Modus Ponens in (1), are processed on-line (see Lea, 1995). This has carried over to the neuroimaging literature, where conditionals are one of the main connectives used to determine the neural centers of logical processing (see Monti et al., 2009; Noveck et al., 2004; Prado et al., 2010; Reverberi et al., 2009). Other approaches to conditionals come from Mental Models (Johnson-Laird & Byrne, 2002) as well as the “new paradigm” (Baratgin et al., 2015), which couches conditional reasoning in probability.

7. O’Brien et al. (1989) argued that children do not appreciate the quantified aspects of If-then sentences. When given a sentence of the sort “All the boxes that have a dog have an apple,” performance was much more adult-like for the 7-year-olds. In this respect, children’s performance here does share something with scalar terms in that a test question such as “Some of the boxes have animals” can be understood as, “There is a box with an animal” or “There are at least two boxes with animals.”

8. One of my early lessons as an experimental pragmatist came from Larry Horn who told me that I should not so readily place all neo-Griceans into one basket because, he said, he himself did not make the claim that scalar enrichments should be considered effortless default inferences the way Levinson did. It is with that in mind when I say, “As far as I know.” If someone did make this proposal and I failed to find it, I apologize in advance.

9. The authors did not report statistical tests, so this point is based on eyeballing the presented data. For other data that point in the same direction, the reader is directed to van der Henst et al. (2006), to be described later.

10. As Jean-Baptiste Van der Henst points out, it could be that no one has considered testing Theory of Mind abilities with respect to conditional reasoning performance. That said, it appears that the two sorts of phenomena afford different kinds of studies on individual differences.

11. This is akin to the labels Bott and Noveck (2004) used when participants’ Some-and-perhaps-all reading was called “logical” and their Some-but-not-all “pragmatic,” except that here the Rejectors are “logical” and the Endorsers are arguably “pragmatic.”

12. We also reported a P3b with the arrival of the minor premise of the MP inference, which is an indication that an expectation was satisfied (Bonnefond & van der Henst, 2009).