Two Sides of Modus Ponens

Reuben Stern* Stephan Hartmann†

July 17, 2018

Forthcoming in *The Journal of Philosophy*

Abstract

McGee (1985) argues that it is sometimes reasonable to accept both x and $x \to (y \to z)$ without accepting $y \to z$, and that *modus ponens* is therefore invalid for natural language indicative conditionals. Here, we examine McGee's counterexamples from a Bayesian perspective. We argue that the counterexamples are genuine insofar as the joint acceptance of x and $x \to (y \to z)$ at time t does not generally imply constraints on the acceptability of $y \to z$ at t, but we use the distance-based approach to Bayesian learning to show that applications of *modus ponens* are nevertheless guaranteed to be successful in an important sense. Roughly, if an agent *becomes* convinced of the premises of a *modus ponens* argument, then she should likewise *become* convinced of the argument's conclusion. Thus we take McGee's counterexamples to disentangle and reveal two distinct ways in which arguments can convince. Any general theory of argumentation must take stock of both.

*Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – https://sites.google.com/view/reubenstern/home – reuben.stern@gmail.com.

†Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – http://www.stephanhartmann.org – s.hartmann@lmu.de.

1 Introduction

McGee (1985) argues that *modus ponens* is invalid for natural language indicative conditionals on the grounds that it is sometimes reasonable to accept x and $x \rightarrow (y \rightarrow z)$ without accepting $y \rightarrow z$. Though it is controversial whether McGee successfully invalidates *modus ponens* by showing that x and $x \rightarrow (y \rightarrow z)$ can be *true* while $y \rightarrow z$ is *false*,[1] McGee does show that one can reasonably accept x and $x \rightarrow (y \rightarrow z)$ while not accepting $y \rightarrow z$. Consider McGee's (1985, p. 463) example:

(1) If that animal is a fish, then if it has lungs, it's a lungfish.

(2) That animal is a fish.

(3) If that animal has lungs, then it's a lungfish.

Just because you think (2) that the animal is a fish and (1) that if it's a fish, then if it has lungs, it's a lungfish, doesn't mean you should think (3) that if the animal has lungs, then it's a lungfish. Indeed, it seems reasonable to conjecture that if it has lungs, it's not a fish at all. So it seems reasonable to accept (1) and (2) while judging (3) to be utterly unacceptable.

The same is not true for standard applications of *modus ponens* not involving "right-nested" conditionals, i.e., conditionals whose consequents are themselves conditionals. Consider:

(4) If that animal is a fish, then it has gills.

(5) That animal is a fish.

(6) That animal has gills.

In the event that you accept (4) and (5), it seems that you should accept (6) – or at least that your opinion about (6) can't wildly diverge from your opinion about (4) and (5) in the way that your opinion about (3) can wildly diverge from your opinions about (1) and (2). So it seems that there is some intuitive sense in which the convincingness of *modus ponens* arguments whose major premises include right-nested conditionals comes apart from the convincingness of standard *modus ponens* arguments.

In this paper, we assess these arguments from a Bayesian perspective and argue that there is one sense in which *modus ponens* arguments are bound to convince no matter whether they involve right-nested conditionals, and another in which only standard *modus ponens* arguments are convincing. First, we use synchronic probability theory to show that arguments involving right-nested conditionals differ from standard *modus ponens* arguments insofar as acceptance of their premises at time t does not generally constrain the set of rational attitudes that can be

[1] See Bledin (2015), Piller (1996), and Sinnott-Armstrong et al. (1986).

adopted towards their conclusions at t, while the same is not true of standard *modus ponens* arguments. Then, we use the distance-based approach to Bayesian learning to show that all *modus ponens* arguments are unified in the sense that they are *diachronically* convincing.[2] That is, all *modus ponens* arguments are convincing (roughly) insofar as it is rational to *become* convinced of their conclusions upon *becoming* convinced of their premises.[3]

This means that McGee's "counterexamples" are interesting not only insofar as they show (or fail to show) that the premises of a *modus ponens* argument can be true while the conclusion is false, but also because they drive a wedge between the distinct ways in which arguments can convince. On the one hand, these arguments (like standard *modus ponens* arguments) are convincing in the sense that it is rational to become more convinced of their conclusions when you learn their premises. On the other hand, unlike standard *modus ponens* arguments, these arguments are *not* convincing in the sense that your standing attitudes towards their premises constrain what you should think (now) about their conclusions. Thus we can draw a lesson from McGee's "counterexamples" that has not yet been drawn in the literature: there are at least two distinct ways in which arguments can qualify as convincing, and any general theory of argumentation should take stock of both.

2 The Synchronic Bayesian Perspective

> If two people are arguing, If p, will q? and both are in doubt as to p, they are adding p hypothetically to their stock of knowledge, and arguing on that basis about q; they are fixing their degrees of belief in q given p. (Ramsey, 1929)

Ramsey's words have inspired a great philosophical tradition of treating the acceptability of a natural language indicative conditional as the probability of its consequent given its antecedent. This is most conspicuously stated by Adams (1975), who says that the following equation holds for all simple conditionals, $x \to y$, such that $P(x) > 0$:[4]

Adams' Acceptability Thesis (AAT): $Acc(x \to y) = P(y|x)$.

AAT has struck many philosophers as intuitive and has been shown time and

[2]We introduce the distance-based approach to Bayesian learning in Section 3.

[3]We state the sense in which all *modus ponens* arguments are convincing more precisely in Section 3.

[4]Adams initially expressed his thesis in terms of probability rather than acceptability. But because he explicitly warned against interpreting the probability of the conditional as the probability of its truth, and instead interpreted the probability of the conditional as the assertability or acceptability of the conditional (à la Ramsey), AAT can be stated as we state it here.

time again to capture actual human reasoning about conditionals.[5] Various triviality results (e.g., Lewis 1976 and Hájek 1989) yield some reason to doubt that AAT can be maintained when $x \rightarrow y$ is regarded as a proposition, but there are some (e.g., Bennett 2003 and Edgington 1995) who are happy to maintain that indicative conditionals are not propositions, and still others (e.g., Douven 2016) who argue that indicative conditionals can be propositional even given AAT and the triviality results.

Anyway, our concern here is not with whether conditionals qualify as propositions, but is rather with how the acceptance of conditionals can be plausibly modeled (since we are ultimately concerned with the implications of accepting conditionals in the context of *modus ponens* arguments). In this setting, it seems reasonable to assume AAT simply because, as has been empirically demonstrated, most of us adopt a tendency to assign degrees of belief to y under the supposition that x to the extent that we accept $x \rightarrow y$.[6] This means that we can represent the attitudes that one has towards the premises and conclusion of a standard *modus ponens* arguments in the following terms:

(a) $P(y|x)$;

(b) $P(x)$;

(c) $P(y)$.

This doesn't quite get us what we need in order to assess *modus ponens* arguments involving right-nested conditionals. The reason is that the applicability of AAT is limited to simple conditionals, and right-nested conditionals are not simple conditionals. Consider $x \rightarrow (y \rightarrow z)$. If we were to apply AAT to this conditional, it would seem that $Acc(x \rightarrow (y \rightarrow z)) = P((z|y)|x)$, but there is no such probability expression as $P((z|y)|x)$.

In order to apply AAT to right-nested conditionals, we require some way of understanding $x \rightarrow (y \rightarrow z)$ in terms that can readily be expressed by the probability calculus. Luckily, there is one. In his original (1985) paper, McGee argues

[5]In recent years, it has been shown that there may be some contexts in which AAT does not capture human reasoning about conditionals. For example, Douven and Verbrugge (2013) show that AAT may get things wrong when x and y are completely independent and unrelated. Nevertheless, there is a wide body of literature (referenced in Douven and Verbrugge 2013) showing that AAT gets things right in many contexts, and there is no reason to believe that any of the contexts that we consider here are exceptions to this rule (since, e.g., we deal only with conditionals whose antecedents are relevant to their consequents).

[6]To some, it may sound as though we are committing the is-ought fallacy because we use empirical results to draw conclusions about the *normative* concept of 'acceptability'. Though we cannot settle this thorny issue here, we believe that we are justified in using empirical results to draw conclusions about the conditions under which sentences are acceptable because the relevant norms derive from linguistic convention, and this is precisely what is being empirically probed.

that we are equally willing to accept $x \rightarrow (y \rightarrow z)$ and $(x \wedge y) \rightarrow z$.[7] For example, it seems that we are just as willing to accept"if that animal is a fish, then if it has lungs, it's a lungfish" as we are willing to accept "if that animal is a fish and has lungs, then it's a lungfish." Following McGee, then, we can posit "Acceptability Import-Export" (AIE):

$$\text{AIE: } Acc(x \rightarrow (y \rightarrow z)) = Acc((x \wedge y) \rightarrow z).$$

Together with AAT, AIE implies that $Acc(x \rightarrow (y \rightarrow z)) = P(z|x \wedge y)$.[8] This is a nice result for anyone interested in analyzing the acceptability of indicative conditionals in terms of conditional probabilities (since it extends the analysis to right-nested conditionals), but one can reasonably wonder whether AIE accurately captures actual reasoning about nested conditionals. As things turn out, van Wijnbergen-Huitink et al. (2015) find no significant difference between the extent to which humans believe (or accept) the right-nested conditional and the extent to which humans believe (or accept) its imported form. So it seems that AIE benefits from empirical confirmation, and that it is therefore plausible to understand $Acc(x \rightarrow (y \rightarrow z))$ in terms of $P(z|x \wedge y)$.[9] This means that we can represent the attitudes that one has towards the premises and conclusion of an instance of *modus ponens* involving a right-nested conditional as follows:

(d) $P(z|x \wedge y)$;

(e) $P(x)$;

(f) $P(z|y)$.

From the synchronic Bayesian perspective, the constraints implied on (c) by (a) and (b) are very different from the constraints implied on (f) by (d) and (e). In fact, while the values of (a) and (b) fix a lower bounds for (c), almost no constraint is implied by the values of (d) and (e) on (f).[10]

[7]This is closely related to the principle, *import-export*, according to which $x \rightarrow (y \rightarrow z)$ and $(x \wedge y) \rightarrow z$ are logically equivalent. McGee proves that any connective, $\rightarrow$, that satisfies some basic constraints and validates both import-export and *modus ponens* must be logically equivalent to the material conditional. Thus McGee shows that we must either (i) analyze the natural language conditional as the material conditional, (ii) deny the validity of *modus ponens*, (iii) deny the validity of import-export, or (iv) analyze the conditional such that it fails to satisfy McGee's basic constraints. See McGee (1985, pp. 465-466) for the details.

[8]As an anonymous referee helpfully points out, this follows only when AIE is restricted to settings where $P(x \wedge y) > 0$ (since AAT applies only in these settings).

[9]Here again, one might worry that we are guilty of committing the is-ought fallacy. Our response is the same as earlier. Because norms of acceptability derive from linguistic convention, norms of acceptability can be confirmed by empirical results about linguistic convention.

[10]The reason for the inclusion of 'almost' is that (d) and (e) do imply constraints on (f) in the special case where $P(x) = 1$ (since $P(z|x \wedge y)$ reduces to $P(z|y)$ when $P(x) = 1$). But in the great many contexts where (d) and (e) are non-extreme, they imply no lower bounds for (f).

This is easy to see when we use probability theory to expand (c) and (f). If we apply the law of total probability to (c), we discover that:

$$P(y) = P(y|x)\,P(x) + P(y|\neg x)\,P(\neg x).$$

Since the first two probabilities correspond to (a) and (b), it is clear that the acceptability of (c) must be at least as great as the product of the acceptabilities of (a) and (b).[11] This means that if we accept (a) and (b) to a high degree, then we must likewise accept (c) to a reasonably high degree. Of course we might coherently take (c) to be slightly less acceptable than either (a) or (b) (since the product of (a) and (b) is less than (a) and less than (b) when neither (a) nor (b) equals 1), but (c) cannot dip too far below the acceptability of the premises when both (a) and (b) are reasonably high since, for example, (c) must be at least .81 when (a) is at least .9 and (b) is at least .9.

We cannot straightforwardly apply the law of total probability to (f) because it expresses a conditional probability, but with a little bit of algebraic manipulation, we arrive at the following expansion of (f):

$$P(z|y) = P(z|x \wedge y)\,P(x|y) + P(z|\neg x \wedge y)\,P(\neg x|y).$$

Though the first probability in the expansion corresponds to (d), none of the other probabilities appear in the premises. This means that you can coherently assign (f) a probability as low as 0 (or as high as 1) even when you regard (d) and (e) as highly acceptable.[12] For example, if you assign .99 to (d) and .99 to (e), you can coherently judge (f) to be utterly unacceptable because you can coherently assign the last three probabilities of the expansion extremely low values, and thereby derive an extremely low probability for (f).

From the synchronic Bayesian perspective, then, it makes perfect sense to find it utterly unacceptable that (3) if the animal has lungs, then it's a lungfish, even when you find it highly acceptable that (1) if it's a fish, then if it has lungs, it's a lungfish, and that (2) it's a fish. The same pattern does not hold for standard applications of *modus ponens*. If you find it highly acceptable that (4) if it is a fish, then it has gills, and that (5) the animal is a fish, then – on pain of incoherence – you must find it reasonably acceptable that (6) it has gills. The synchronic Bayesian perspective thus appears to vindicate the initial thought that the convincingness of *modus ponens* arguments whose major premises include right-nested conditionals comes apart from the convincingness of standard *modus ponens* arguments.

[11] This is because the product of the second two probabilities cannot be negative.

[12] Again, if you regard (e) as *fully* acceptable – i.e., if $Acc(x) = 1$ – then you must regard (f) and (e) as equally acceptable. But if you have any doubts at all about (d) and (e), then your opinions about (d) and (e) do not imply constraints on your opinion of (f). This underscores an important point. If you *fully* accept the minor premise of McGee's example, then you must regard its major premise and conclusion as equally acceptable. This means that in the special case where the agent fully accepts the minor premise of McGee's putative counterexample, *modus ponens* actually does qualify as *synchronically* convincing in an important sense.

3 The Diachronic Bayesian Perspective

At this juncture, it may seem as though *modus ponens* arguments involving right-nested conditionals are unsuccessful when viewed from the Bayesian lens. If you accept their premises, you can think whatever you want about their conclusions. So they are not convincing arguments.

But this is not the whole Bayesian story. Though it is true that your standing joint acceptance of x and $x \to (y \to z)$ at time t does not imply that you should at all accept $y \to z$ at t, there may be some other Bayesian sense in which these arguments qualify as convincing.[13] In a recent paper, Eva and Hartmann (forthcoming) consider whether and when various argument forms are convincing inasmuch as it is rational to *become* convinced of their conclusions upon *learning* their premises. That is, rather than asking whether an agent's current attitudes towards the premises of an argument commit her to having certain attitudes towards its conclusion, they ask how and whether an agent's attitudes towards an argument's conclusion should *change* upon becoming convinced of its premises.

In order to probe questions of this sort, Eva and Hartmann ('EH') deploy the distance-based approach to Bayesian learning, according to which we can determine how an agent should update her credal state after learning some new information by using a distance or divergence measure to determine which of the many probability functions compatible with what she learned is *closest* to her initial (prior) probability function. Thus when the agent learns to fully accept both that the animal has gills if it is a fish and that the animal is a fish, we determine what the closest probability function Q is to her prior probability function P that obeys the constraints that $Q(gills|fish) = 1$ and that $Q(fish) = 1$. The agent should then adopt this probability function because doing so yields the maximally conservative update – or, put differently, because adopting this probability function embodies the minimal revision that manages to incorporate everything learned by the agent across the update.

Which probability function counts as closest to the prior depends on what distance or divergence metric is used to calculate closeness. Eva and Hartmann focus primarily on the minimization of *Kullback-Leibler divergence* ('KL'). The KL divergence is defined as follows. Let $\mathrm{S}_1, \ldots, \mathrm{S}_\mathrm{n}$ be the possible values of a random variable S over which the probability distributions P and Q are defined. Then

$$D_{KL}(Q||P) := \sum_{i=1}^{n} q_i \cdot \log\left(q_i/p_i\right),$$

where we have used the abbreviations $p_i := P(\mathrm{S_i})$ and $q_i := Q(\mathrm{S_i})$.[14]

[13] An agent plausibly accepts a conditional when she judges its acceptability to be above a reasonable (perhaps contextually determined) threshold.

[14] Note that the KL divergence is not symmetrical and that it may not satisfy the triangle inequality. This means that it is not a distance measure in the mathematical sense of the term.

Like EH, we focus on minimizing KL because it yields Bayesian conditionalization when the agent learns something with certainty, and Jeffrey conditionalization when the agent acquires a new non-extreme probability estimate for something.[15] Where KL minimization earns its keep over these more standard updating procedures is when the agent learns something that cannot be straightforwardly represented in terms of conditionalization (or Jeffrey conditionalization) – e.g., when the agent learns a new *conditional* probability estimate, or when the agent learns to treat some propositions as probabilistically (in)dependent. This is of utmost importance when determining whether an agent should accept the conclusion of an argument upon learning its premises for the simple reason that the premises of arguments often contain conditionals, and it is plausible to represent learning $x \to y$ in terms of an increase in the agent's subjective conditional probability of y given x.

EH use the distance-based approach to Bayesian learning to prove results about the value of deductively valid arguments from the diachronic Bayesian perspective. For example, under the assumption of AAT, EH show that deductively valid argument forms like *modus ponens* and *modus tollens* differ from so-called "fallacious" argument forms like *affirming the consequent* and *denying the antecedent* insofar as agents who learn the deductively valid arguments' premises should, as a matter of necessity, become more convinced of their conclusions, while agents who learn the other arguments' premises need not become more convinced of their conclusions.[16,17]

Of special interest here are EH's two results about *modus ponens*. First, EH prove that if an agent becomes fully convinced of $x \to y$ and becomes more convinced of x, then, when the agent minimizes KL, she necessarily becomes more confident in y. Second, EH prove that if an agent becomes more convinced of $x \to y$ and does not alter her degree of belief in y across some update, then, when the agent minimizes KL, she necessarily becomes more confident in y.[18] The case

[15]KL is actually a member of a broader family of f-divergences that all yield Bayesian conditionalization and Jeffrey conditionalization. EH do not offer any principled reason to prefer KL to other f-divergences, but work with KL because it is well-known and used by other Bayesians. Like EH, we are open to deploying other f-divergences, but work primarily with KL for similar reasons. It is worth mentioning, though, that KL is the only f-divergence that is a Bregmann divergence, and that it can therefore be argued for on the grounds that it is the only f-divergence that defines updates that minimize expected inaccuracy when measured by strictly proper scoring rules. (See Amari (2009) and Pettigrew (2016) for background.) At any rate, as we report our own results, we keep tabs (in the footnotes) of which results hold for KL and which results hold for the entire class of f-divergences.

[16]Hahn and Oaksford (2006) use Bayesian tools to argue that these "fallacious" argument forms are actually legitimate modes of argumentation in many contexts.

[17]Eva and Hartmann show this by assessing whether the probability of the conclusion must increase when the agent minimizes KL – no matter her prior – while satisfying certain constraints corresponding to what she learns about the premises. The precise content of these constraints varies with the particular result. In the next paragraph, we outline the precise constraints where the EH results apply in the case of standard instances of *modus ponens*.

[18]More formally, if P is the prior and Q is the posterior, EH prove, first, that $Q(y) > P(y)$ if

where an agent becomes fully convinced of both premises is the special case of the first result where the increase in probability of the minor premise goes all the way to certainty.[19]

If the EH results extended to applications of *modus ponens* involving right-nested conditionals, then there would already be proof that the McGee cases exemplify an argument form where *modus ponens* is convincing diachronically (in the sense specified by the EH results), but not synchronically (because of the findings from the last section). But the EH results are limited to contexts where x and y are Boolean combinations of atomic propositions, and it is therefore still unclear how applications of *modus ponens* involving right-nested conditionals fare when assessed from the diachronic Bayesian perspective.[20]

In order to probe this question, we can model the acceptability of these premises in terms of (d) and (e), and the acceptability of the conclusion in terms of (f). Then we can gauge what happens when we minimize KL while satisfying constraints like those deployed in the EH results about standard applications of *modus ponens*.

We assume that the agent has a prior probability distribution P over the binary propositional variables X, Y, Z (with the values x and $\neg x$ etc.). She then learns the premises of the argument. They result in constraints on the posterior probability distribution Q. Let us now consider the scenarios that EH consider in the context of standard applications of *modus ponens*.

In the first scenario, we consider the situation where the acceptability of the major premise, $x \rightarrow (y \rightarrow z)$, goes to 1 and the acceptability of the minor premise, x, increases. Then we ask whether the acceptability of the conclusion, $y \rightarrow z$, must increase. Given AAT and AIE, this leads to the following two constraints on the posterior:

Premise MP1: $Q(z|x \wedge y) = 1$.

Premise MP2: $Q(x) > P(x)$.

We can then prove the following theorem (all proofs are in the appendix):

$Q(y|x) = 1$ and $Q(x) > P(x)$, and, second, that $Q(y) > P(y)$ if $Q(y|x) > P(y|x)$ and $Q(x) = P(x)$.

[19]It is perhaps natural to wonder whether the acceptability of the conclusion goes up whenever one comes to accept both premises more than she used to. As EH (forthcoming) note, this result interestingly does not hold – basically because the acceptability of the conclusion can decrease when the acceptability of both premises increase, provided that the minor premise and the conclusion are heavily anti-correlated in the prior.

[20]Throughout this paper, it should be clear that we are assuming (with Eva and Hartmann) that the indicative conditional is not the material conditional. Were the indicative conditional the material conditional, then AAT would not be plausible. It is reasonable to wonder whether our treatment of indicative conditionals requires that we assume anything else about the semantics of the indicative conditional. For all we know, the answer may be yes, but we would like to leave this issue for later. It is notoriously difficult to develop a semantics for indicative conditionals that vindicates AAT and AIE while treating indicative conditionals as propositions, and we remain open to the possibility that indicative conditionals are not propositions. But we also remain open to the possibility that indicative conditionals are, in fact, propositions, and that AAT and AIE can be vindicated by pragmatics.

Theorem 1 *An agent considers the propositional variables X, Y, and Z and has a prior probability distribution P defined over them. If the agent minimizes any f-divergence (including KL) between the posterior probability distribution Q and P while satisfying the constraints in* **MP1** *and* **MP2**, *then the acceptability of $y \to z$ increases, i.e., $Q(z|y) > P(z|y)$.*

In the second scenario, we consider the situation where the acceptability of the major premise, $x \to (y \to z)$, increases to some value smaller than 1, and the probability of the minor premise, x, does not change. This leads to the following two constraints on the posterior:

Premise MP1': $Q(z|x \wedge y) > P(z|x, y)$.

Premise MP2': $Q(x) = P(x)$.

We can then prove the following theorem:

Theorem 2 *An agent considers the propositional variables X, Y, and Z, and has a prior probability distribution P defined over them. If the agent minimizes KL between the posterior probability distribution Q and P while satisfying the constraints in* **MP1′** *and* **MP2′** *then the acceptability of $y \to z$ increases, i.e., $Q(z|y) > P(z|y)$.*

These two theorems show that applications of *modus ponens* involving right-nested conditionals are *diachronically* convincing in the very same way that standard applications of *modus ponens* are, even though applications of *modus ponens* involving right-nested conditionals are different from standard applications inasmuch as they are not at all *synchronically* convincing. That is, even though it is true that your standing joint acceptance of x and $x \to (y \to z)$ at t does not generally constrain your opinion of $y \to z$ at t, you should become *more* convinced of $y \to z$ upon either (i) becoming fully convinced of $x \to (y \to z)$ and becoming more convinced of x, or (ii) becoming more convinced of $x \to (y \to z)$ while maintaining your previous opinion of x.[21]

4 Revisiting McGee's Example

In light of these results, it is worth considering McGee's lungfish example once more. Though we have already argued that the synchronic Bayesian perspective vindicates McGee's insight that it is reasonable to accept the premises while regarding the conclusion as utterly unacceptable, it now seems that there is another sense in which we should find the lungfish example convincing. Namely, you should become *more* convinced that if the animal has lungs, then it is a lungfish upon

[21] Throughout this paper, we say that the agent becomes more convinced of a conditional when she comes to judge it as more acceptable than she used to, and say that the agent becomes fully convinced of the conditional when she comes to judge it as fully acceptable (or, equivalently, comes to assign it an acceptability of 1).

either (i) becoming fully convinced that if it's a fish, then if it has lungs, it's a lungfish, and becoming more convinced that it's a fish, or (ii) becoming more convinced that if it's a fish, then if it has lungs, it's a lungfish while remaining equally convinced that it's a fish. Are these results intuitive?

To gauge the intuitiveness of the first result, consider the following story.

Suppose that Matthias goes to the beach with his marine biologist friend, Klara, and that Matthias and Klara are wondering what some animal in the water is. Klara tells Matthias (who knows almost nothing about fish), first, that if that animal is a fish, then if it has lungs, it's a lungfish, and, second, that it's very probably a fish. Because Matthias trusts Klara, he becomes more confident that it's a fish and learns to fully accept that if it is a fish, then if it has lungs, it's a lungfish. Should Matthias become more inclined to accept that it's a lungfish if it has lungs than he used to be?

Intuitively, yes! Because Matthias is now more confident that it's a fish and fully convinced of Klara's claim about fish with lungs, it seems that Matthias should be more inclined to accept that it's a lungfish if it has lungs. After all, when Matthias learns that it's very probably a fish, he seems to acquire reason to think that it's a lungfish if it has lungs, even if he still has reason to regard it as unacceptable (overall) that it's a lungfish if it has lungs. So it seems that the lungfish argument is intuitively convincing in the sense that Matthias should regard its conclusion as *more* acceptable than he used to, even if Matthias need not regard its conclusion as *highly* or *reasonably* acceptable.

Now consider a second story in order to gauge the intuitiveness of the second result.

Suppose that Klara's marine biologist friend, Jacques, joins Matthias and Klara at the beach, and that Klara and Jacques begin to discuss what kind of animal it might be. Klara reminds Jacques that if it's a fish, then if it has lungs, it's a lungfish, but Jacques is not entirely convinced – viz., Jacques is more inclined to accept Klara's claim than he was prior to Klara's reminder (because he needed the reminder), but is not *totally* willing to accept Klara's claim (perhaps because he thinks she may be forgetting about some species of fish with lungs). Since Jacques is a marine biologist himself, he trusts his own estimate that the animal is a fish and doesn't budge upon hearing anything from Matthias or Klara. Should Jacques become more inclined to accept that if the animal has lungs, then it's a lungfish?

Again, the answer seems to be yes. Because Jacques is now more willing to accept Klara's claim about lungfish and just as confident as before that it's a fish, it seems that Jacques should be more confident that it's a lungfish in the event that it has lungs. And since this confidence corresponds to the acceptability of the claim that it's a lungfish if it has lungs, it seems that Jacques should, in fact, be more willing to accept that the animal is a lungfish if it has lungs after talking things over with Klara, even if he needn't think that it is (overall) acceptable. So here, too, it seems that the lungfish argument is convincing in the sense that Jacques should regard its conclusion as *more* acceptable than he used to, even if he need not regard its conclusion as *highly* or *reasonably* acceptable.

5 Conclusion

Though only standard applications of *modus ponens* prove to be synchronically convincing (in the sense outlined in Section 2), every application of *modus ponens* appears to be *diachronically* convincing (in the sense probed by EH). It should not be surprising that there is some unifying inferential property of *modus ponens* arguments, given their common linguistic structure. But McGee's cases are largely valuable because they drive a wedge between two ways that arguments can convince. That is, though diachronic convincingness and synchronic convincingness are a package deal in standard cases of *modus ponens*, they are decoupled in cases involving right-nested conditionals, and therefore revealed to be distinct. So it seems that by reflecting on McGee's cases, two distinct senses in which arguments can be convincing are revealed, and that any general theory of argumentation must take stock of both.

Acknowledgements

We thanks Benjamin Eva, Malcolm Forster, Lina Lissia, Mike Oaksford, David Over, Jan Sprenger, Olav Vassend, Jon Williamson and two anonymous referees for helpful feedback.

A Proofs

We represent the probability distributions P and Q in Table 1 with "1" representing x and "0" representing $\neg x$ etc.

Table 1: The probability distributions P and Q

X	Y	Z	P	Q
1	1	1	p_1	q_1
1	1	0	p_2	q_2
1	0	1	p_3	q_3
1	0	0	p_4	q_4
0	1	1	p_5	q_5
0	1	0	p_6	q_6
0	0	1	p_7	q_7
0	0	0	p_8	q_8

A.1 Theorem 1

We consider the situation where the probability of the conditional goes to 1 and the probability of the minor premise increases. This leads to the following constraints:

1. $Q(z|x \wedge y) = 1$. Hence $Q(x \wedge y \wedge z) = Q(x \wedge y)$. Hence $Q(x \wedge y \wedge \neg z) = q_2 = 0$.
2. $Q(x) > P(x)$.

The second constraint amounts to

$$q_1 + q_3 + q_4 - p_1 - p_2 - p_3 - p_4 - \delta = 0\,, \tag{1}$$

with $\delta > 0$. Note that this constraint implies that

$$a := p_1 + p_2 + p_3 + p_4 < 1. \tag{2}$$

As a further constraint, we make sure that

$$q_1 + q_3 + \cdots + q_8 - 1 = 0. \tag{3}$$

Hence we have to minimize

$$L = \sum_{i \neq 2} p_i\, f(q_i/p_i) + \lambda\,(q_1 + q_3 + q_4 - a - \delta) + \mu\,(q_1 + q_3 + \cdots + q_8 - 1),$$

where f is some convex function with $f(1) = 0$ ("f-divergence"). To do so, we compute the following derivatives:

$$\begin{aligned} \frac{\partial L}{\partial q_l} &= f'(q_l/p_l) + \lambda + \mu = 0, \quad \text{for } l = 1, 3, 4 \\ \frac{\partial L}{\partial q_m} &= f'(q_m/p_m) + \mu = 0, \quad \text{for } m = 5, 6, 7, 8 \end{aligned}$$

Setting them equal to zero yields

$$q_l = \alpha\, p_l \quad , \quad q_m = \beta\, p_m, \tag{4}$$

with two parameters α and β which have to be fixed to make sure that the constraints (1) and (3) are satisfied. To determine them, we insert the first equation in eqs. (4) into eq. (1) and obtain

$$\alpha\,(p_1 + p_3 + p_4) = a + \delta. \tag{5}$$

Hence

$$\alpha = 1 + \frac{p_2 + \delta}{p_1 + p_3 + p_4}. \tag{6}$$

(Note that $p_1 > 0$ to make sure that the first constraint can be satisfied.) Next, we insert eqs. (4) into eq. (3) and obtain:

$$\alpha\,(p_1 + p_3 + p_4) + \beta\,(p_5 + p_6 + p_7 + p_8) = 1 \tag{7}$$

Using eq. (5), we obtain

$$a + \delta + \beta\,(p_5 + p_6 + p_7 + p_8) = 1. \tag{8}$$

Using the fact that $p_1 + p_2 + \cdots + p_8 = 1$ and setting $\overline{a} := 1 - a$, we obtain

$$a + \delta + \beta\,\overline{a} = 1. \tag{9}$$

Hence,

$$\beta = 1 - \frac{\delta}{\overline{a}}. \tag{10}$$

We also express α in terms of the variable a:

$$\alpha = 1 + \frac{p_2 + \delta}{a - p_2} \tag{11}$$

Next, we calculate

$$\begin{aligned} P(z|y) &= \frac{P(y \wedge z)}{P(y)} = \frac{p_1 + p_5}{p_1 + p_2 + p_5 + p_6} \\ Q(z|y) &= \frac{Q(y \wedge z)}{P(y)} = \frac{q_1 + q_5}{q_1 + q_5 + q_6} = \frac{\alpha\,p_1 + \beta\,p_5}{\alpha\,p_1 + \beta\,(p_5 + p_6)}. \end{aligned}$$

Hence,

$$\Delta = Q(z|y) - P(z|y) = \frac{\Delta'}{(p_1 + p_2 + p_5 + p_6)\,(\alpha\,p_1 + \beta\,(p_5 + p_6))},$$

with

$$\begin{aligned} \Delta' &= (\alpha\,p_1 + \beta\,p_5)(p_1 + p_2 + p_5 + p_6) - (p_1 + p_5)(\alpha\,p_1 + \beta\,(p_5 + p_6)) \\ &= (\alpha - \beta)\,p_1\,p_6 + \alpha\,p_1\,p_2 + \beta\,p_2\,p_5. \end{aligned}$$

Finally, we note that $\alpha, \beta > 0$ and that

$$\begin{aligned} \alpha - \beta &= \frac{p_2 + \delta}{a - p_2} + \frac{\delta}{\overline{a}} \\ &= \frac{p_2 \cdot \overline{a} + \overline{p_2} \cdot \delta}{(a - p_2)\,\overline{a}}. \end{aligned} \tag{12}$$

From eq. (12) it is easy to see that $\alpha - \beta > 0$ if $\delta > 0$. Hence, $\Delta > 0$ if $\delta > 0$, which is what we wanted to show. ∎

A.2 Theorem 2

We consider the situation where the probability of the conditional increases and the probability of the minor premise does not change. That is, we request that

1. $Q(z|x \wedge y) > P(z|x \wedge y)$. This amounts to:

$$\overline{\alpha}\, p_2\, q_1 - (p_1 + \alpha\, p_2)\, q_2 = 0, \tag{13}$$

with $0 < \alpha \leq 1$. This can be seen as follows. Let $Q(z|x \wedge y) = P(z|x \wedge y) + \delta \leq 1$ with $\delta > 0$. This implies that $\delta \leq 1 - P(z|x \wedge y) = P(\neg z|x \wedge y)$. We therefore set

$$\delta = \alpha \cdot P(\neg z|x \wedge y) = \frac{\alpha\, p_2}{p_1 + p_2}$$

with $0 < \alpha \leq 1$. The parameter α regulates how much the conditional probability increases. $\alpha = 0$ means no increase at all, $\alpha = 1$ is the maximal increase. In this case the new probability $Q(z|x \wedge y)$ equals 1. Hence,

$$\begin{aligned} Q(z|x \wedge y) &= \frac{q_1}{q_1 + q_2} \\ &= P(z|x \wedge y) + \frac{\alpha\, p_2}{p_1 + p_2} \\ &= \frac{p_1 + \alpha\, p_2}{p_1 + p_2}. \end{aligned}$$

After some algebra, we obtain the constraint (13). Note also that the first condition requires that

$$0 < p_1, p_2 < 1. \tag{14}$$

2. $Q(x) = P(x)$. Hence (using eq. (2)),

$$q_1 + q_2 + q_3 + q_4 - a = 0. \tag{15}$$

3. All q_i sum up to 1. Hence,

$$q_1 + \cdots + q_8 - 1 = 0. \tag{16}$$

Adding these constraints via Lagrange multipliers to the KL-divergence, we arrive at the following function which we have to minimize:

$$\begin{aligned} KL &= \sum_{i=1}^{8} q_i \log \frac{q_i}{p_i} + \lambda \left(\overline{\alpha}\, p_2\, q_1 - (p_1 + \alpha\, p_2)\, q_2\right) \\ &+ \mu \left(q_1 + \cdots + q_4 - a\right) + \nu \left(q_1 + \cdots + q_8 - 1\right) \end{aligned} \tag{17}$$

To find the minimum, we differentiate:

$$\begin{aligned} \frac{\partial KL}{\partial q_1} &= 1 + \log \frac{q_1}{p_1} + \lambda\, \overline{\alpha}\, p_2 + \mu + \nu \\ \frac{\partial KL}{\partial q_2} &= 1 + \log \frac{q_2}{p_2} - \lambda \left(p_1 + \alpha\, p_2\right) + \mu + \nu \\ \frac{\partial KL}{\partial q_3} &= 1 + \log \frac{q_3}{p_3} + \mu + \nu \\ \frac{\partial KL}{\partial q_4} &= 1 + \log \frac{q_4}{p_4} + \mu + \nu \\ \frac{\partial KL}{\partial q_m} &= 1 + \log \frac{q_m}{p_m} + \nu, \quad m = 5, 6, 7, 8 \end{aligned}$$

Setting these expressions equal to zero, we obtain:

$$q_1 = u\, v\, \mathrm{e}^{-\lambda\, \overline{\alpha}\, p_2}\, p_1 \quad , \quad q_2 = u\, v\, \mathrm{e}^{\lambda\, (p_1 + \alpha\, p_2)}\, p_2 \tag{18}$$

$$q_3 = u\, v\, p_3 \quad , \quad q_4 = u\, v\, p_4 \tag{19}$$

$$q_m = u\, p_m \tag{20}$$

with $u := \mathrm{e}^{-1-\nu}$ and $v := \mathrm{e}^{-\mu}$.

Using eqs. (15), (16) and (20) and the fact that the p_i sum up to 1, we obtain: $a + \overline{a}\, u = 1$. Hence,

$$u = 1. \tag{21}$$

Inserting eqs. (18) into eq. (13) yields:

$$\mathrm{e}^{\lambda} = \left(\frac{\overline{\alpha}\, p_1}{p_1 + \alpha\, p_2}\right)^{\frac{1}{p_1 + p_2}} \tag{22}$$

Inserting eqs. (18) and (19) into eq. (15) then yields:

$$v = \frac{a}{\mathrm{e}^{-\lambda\,\overline{\alpha}\,p_2}\,p_1 + \mathrm{e}^{\lambda\,(p_1+\alpha\,p_2)}\,p_2 + p_3 + p_4} \tag{23}$$

With eqs. (21), (22), (23), the q_i are fully determined.

Let us now explore whether $Q(z|y) > P(z|y)$, i.e. whether

$$\frac{q_1 + q_5}{q_1 + q_2 + q_5 + q_6} > \frac{p_1 + p_5}{p_1 + p_2 + p_5 + p_6}. \tag{24}$$

This condition is equivalent to

$$\chi := \frac{p_2 + p_6}{p_1 + p_5} \cdot \frac{q_1 + q_5}{q_2 + q_6} > 1. \tag{25}$$

Inserting the expressions for q_1, q_2, q_5 and q_6 yields

$$\chi := \frac{p_2 + p_6}{p_1 + p_5} \cdot \frac{a\,p_1 + p_1\,p_5 + f(\alpha)\,p_2\,p_5 + g(\alpha)\,(p_3 + p_4)\,p_5}{p_1\,p_6 + f(\alpha)\,p_2\,(a + p_6) + g(\alpha)\,(p_3 + p_4)\,p_6} \tag{26}$$

with

$$f(\alpha) \;:=\; \frac{\overline{\alpha}\,p_1}{p_1 + \alpha\,p_2} \tag{27}$$

$$g(\alpha) \;:=\; f(\alpha)^{\frac{\overline{\alpha}\,p_2}{p_1+p_2}}\,. \tag{28}$$

These functions have the following properties for $0 < \alpha \leq 1$:

$$0 \leq f(\alpha), g(\alpha) < 1 \tag{29}$$

$$f(\alpha) \leq g(\alpha) \tag{30}$$

Property (30) follows because

$$\frac{\overline{\alpha}\,p_2}{p_1 + p_2} \leq 1.$$

Note further that $\chi > 1$ if the difference Δ between the numerator and the denominator in eq. (26) is greater than zero. We calculate:

$$\Delta = (1 - f(\alpha)) \cdot \Delta_f + (1 - g(\alpha)) \cdot \Delta_g + (g(\alpha) - f(\alpha)) \cdot \Delta_{fg} \tag{31}$$

with

$$\begin{aligned} \Delta_f : &= p_2\,(a\,p_1 + p_1\,p_6 + p_1\,p_5) \\ \Delta_g : &= p_1\,p_6\,(p_3 + p_4) \\ \Delta_{fg} : &= p_2\,p_5\,(p_3 + p_4). \end{aligned}$$

As $\Delta_f > 0$ (see eq. (14)) and $\Delta_g, \Delta_{fg} \geq 0$, we conclude (using properties (29) and (30)) that $\Delta > 0$ (and hence $\chi > 1$). ∎

References

Adams, E. W. (1975): *The Logic of Conditionals*. Dordrecht: Reidel.

Amari, S. I. (2009): 'α-Divergence Is Unique, Belonging to Both f-Divergence and Bregman Divergence Classes', *IEEE Transactions on Information Theory*. 55: 4925–4931.

Bennett, J. (2003): *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Bledin, J. (2015): 'Defending Modus Ponens', *Journal of Philosophy* 112: 57–83.

Douven, I. (2016): *The Epistemology of Indicative Conditionals*. Cambridge: Cambridge University Press.

Douven, I. and S. Verbrugge (2013): 'The Probabilities of Conditionals Revisited', *Cognitive Science* 37: 711–730.

Edgington, D. (1995): 'On Conditionals', *Mind* 104: 235–329.

Eva, B. and S. Hartmann (forthcoming): 'Bayesian Argumentation and the Value of Logical Validity', *Psychological Review*. Preprint available at http://philsci-archive.pitt.edu/14491/.

Hahn, U. and M. Oaksford (2006): 'A Bayesian Approach to Informal Argument Fallacies', *Synthese* 152: 207–236.

Lewis, D. (1976): 'Probabilities of Conditionals and Conditional Probabilities', *Philosophical Review* 85: 297–315.

Hájek, A. (1989): 'Probabilities of Conditionals – Revisited', *Journal of Philosophical Logic* 18: 423–438.

McGee, V. (1985): 'A Counterexample to Modus Ponens', *Journal of Philosophy* 82: 462–471.

Pettigrew, R. (2016): *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Piller, C. (1996): 'Vann McGee's Counterexample to Modus Ponens', *Philosophical Studies* 82: 27–54.

Ramsey, F. P. (1929): 'General Propositions and Causality', in his *Philosophical Papers*, ed. D.H. Mellor, Cambridge: Cambridge University Press, 1990, pp. 145–163.

Sinnott-Armstrong, W., J. Moor, and R. Fogelin (1986): 'Vann McGee's Counterexample to Modus Ponens', *Journal of Philosophy* 83: 296–300.

van Wijnbergen-Huitink, J., S. Elqayam, and D. Over (2015): 'The Probability of Iterated Conditionals', *Cognitive Science* 39: 788–803.