Hypothesis Competition beyond Mutual Exclusivity

Jonah N. Schupbach (University of Utah), jonah.n.schupbach@utah.edu
David H. Glass (Ulster University), dh.glass@ulster.ac.uk

Abstract

Competition between scientific hypotheses is not always a matter of mutual exclusivity. Consistent hypotheses can compete to varying degrees either directly or indirectly via a body of evidence. We motivate and defend a particular account of hypothesis competition by showing how it captures these features. Computer simulations of Bayesian inference are used to highlight the limitations of adopting mutual exclusivity as a simplifying assumption to model scientific reasoning, particularly due to the exclusion of hypotheses that may be true. We end with a case study demonstrating the subtleties involved in hypothesis competition in scientific practice.

1. Introduction

In formal philosophy of science, certain simplifying and idealizing assumptions are often required in order to make questions tractable. The problem is that such assumptions can also get in the way, confounding results derived from our formal models. Results are confounded when they depend upon a part of the model that does not accurately portray the target. Here is a simple, toy example: even an ornate, highly detailed model of Cathédrale Notre-Dame de Paris will not represent every crack, patch, and cranny in the structure's brick walls. To conclude from this model that the actual Cathedral's bricks are somewhat glossy and smooth would be a mistake; an idealized feature of the model will have confounded your "results" should you conclude this.

The nature of the idealizing assumption(s) made by one's formal model naturally depends on the case. However, it is striking how often one particular idealizing assumption shows up (implicitly or explicitly) in formal philosophy of science, namely, the assumption that competing scientific hypotheses are mutually exclusive. This assumption is the subject of our paper.

Many philosophers of science may not think of this as an idealizing assumption. They might think that this feature of a formal model gets the target right—at least often enough for it not to be a worrisome confounder. We accordingly argue in Section 2 that competition is often not a simple matter of mutual exclusivity. If the mutual exclusivity assumption has an appropriate *general* place in formal philosophy of science then, it is as a simplifying assumption allowing questions about competing hypotheses to be tractable, not as an accurate portrayal of reality. Highlighting the various ways in which consistent hypotheses may compete allows us to uncover two desiderata for a fuller and more accurate explication of hypothesis competition. In Section 3, we use these two desiderata to motivate and defend a particular formal explication of hypothesis competition. In Section 4, we use computer simulations and this formal explication to explore the extent to which, and conditions under which, this idealizing assumption can confound work in formal philosophy of science. We focus on the particular case of the study of Bayesian inference and its reliability. In Section 5, we turn briefly to a case study to highlight further the need for a more careful treatment of hypothesis competition in the philosophy of science. We suggest that our explication provides the formal philosopher of science with a far more useful and accurate representation of what it takes for hypotheses to compete in actual scientific practice.

2. Desiderata for Hypothesis Competition

Competition between scientific hypotheses is very often not a matter of mutual exclusivity. It is easy to find actual cases of consistent scientific hypotheses being treated as competitors. There are two distinct ways in which consistent hypotheses may compete, each suggestive of senses in which a satisfactory account of hypothesis competition must part from the simple mutual exclusivity idea.

First, hypotheses that are not strictly speaking mutually exclusive may still do much to directly rebut each other; they may be logically consistent even if they *nearly* rule each other out. To give a simple example, a detective might judge it very unlikely that Smith and Jones carried out the robbery together because he knows they are sworn enemies. The upshot is that hypotheses may compete *to varying degrees* corresponding (at least in this case) to the extent that they disconfirm one another—how close they come to entirely ruling each other out. This gives rise to our first desideratum:

Desideratum 1. Hypothesis competition is a matter of degree.

Second, hypotheses may compete with one another, even in cases where they are not only consistent but perfectly compatible or even supportive of one another. There is no direct conflict between them in such a case, but they may nonetheless compete *indirectly*, via some body of evidence. Hypotheses compete in this way to the extent that adopting either hypothesis undermines the support that the relevant body of evidence provides for the other.

As an illustration of this from science, consider the ongoing debates about the mass extinction at the Cretaceous-Paleogene (K-Pg) boundary which brought an end to the dinosaurs around 66 million years ago. The leading hypothesis is the occurrence at this time in history of a bolide impact (Alvarez et al. 1980; Schulte et al. 2010); but other contending hypotheses include massive volcanism, climate change, and sea level regression. Overall, a vast number of further hypotheses have been proposed (Benton 1990). Cleland (2002, 2011) argues that historical science proceeds by i) the proliferation of rival hypotheses to explain a puzzling body of traces, and ii) a search for a 'smoking gun' to discriminate between them. Many scientists claim that evidence relating to the K-Pg boundary—including the iridium anomaly and the existence of impact ejecta such as shocked quartz and spherules, as well as the discovery of the Chicxulub crater on the Yucatan peninsula in Mexico (Hildebrand et al. 1991) and evidence that ejecta at K-Pg boundary sites show a distribution pattern related to the distance from the crater (Schulte et al. 2010)—provide a smoking gun for the impact hypothesis.

Without considering the indirect pathway to competition, one might legitimately be led to question whether any of these hypotheses are really rivals (or competitors) at all—as Cleland's approach requires. Certainly there is no logical incompatibility between the hypotheses mentioned above; any philosopher of science who explicates competition simply as mutual exclusivity will not be accurately representing this feature of the scientific debate.[1]

[1] It is easy to think of distinct hypotheses, parallel to the above, that *would* be mutually exclusive. For example, instead of simply positing the historic occurrence of a bolide impact, massive volcanic activity, etc., parallel hypotheses might state that *the primary cause of mass extinction at the K-Pg boundary* was bolide impact, or volcanic activity, etc. The primary cause could not be more than one of these events, and so one might argue that competition in this case really does come down to mutual exclusivity. We think this

But note that hypotheses such as impact and volcanism are also not plausibly thought of as direct competitors at all. It is much more reasonable to treat them as independent before any evidence is taken into account since plausibly they do little or nothing to rebut each other directly to any significant degree.[2] Rather, insofar as there is competition between these hypotheses, it comes indirectly by way of the body of evidence for which they both individually claim to account.

The upshot is that a full account of hypothesis competition needs to take into account two different ways in which scientific hypotheses can compete. Figure 1 provides a visual representation of the two distinct "pathways to hypothesis competition." Solid lines with arrows represent logical (deductive or inductive) relationships between propositions. In (a), the absence of an arrow between H_1 and H_2 means that there is no direct support or disconfirming relation between these hypotheses. This means that H_1 and H_2 can only compete indirectly via E as indicated by pathway 1. In (b), H_1 and H_2 have a direct bearing on one another and so they may also compete directly if pathway 2 describes a disconfirming rather than supporting relation.

The foregoing discussion gives rise to our second desideratum:

> **Desideratum 2.** There are two pathways to hypothesis competition: a direct pathway and an indirect pathway via the evidence.

---

argument would be misguided for many reasons, including the following two: (1) So stated, the hypotheses *imply* the occurrence of the explanandum (the mass extinction), and so they all account maximally well for the explanandum in this sense. However, the hypotheses that scientists actually have in mind in this case are evaluated for how well they support the explanandum, some being perceived as more or less capable of accounting for the mass extinction. (2) While these parallel, mutually exclusive hypotheses surely do compete, it remains the case that the original, consistent hypotheses we (and working scientists) have in view can also clearly compete—for one thing, they can do so when they undermine each other's evidential support. By replacing consistent competing hypotheses with parallel, mutually exclusive hypotheses, we are ignoring—not illuminating—the notion of hypothesis competition at work in such cases.

[2] Later we will draw attention to some recent work suggesting a positive dependence.

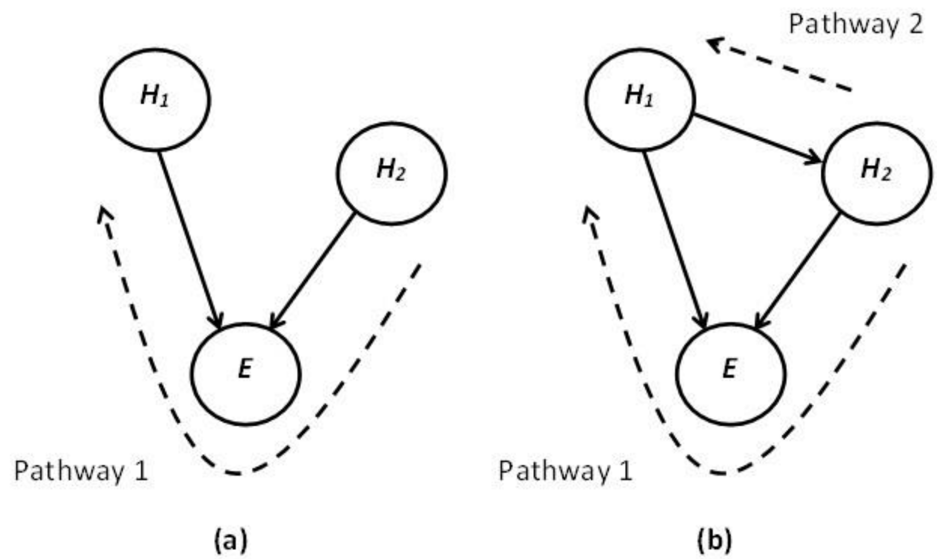

Figure 1. In (a), H_1 and H_2 may only compete indirectly via Pathway 1; in (b), H_1 and H_2 have the potential to compete directly via Pathway 2 or indirectly via Pathway 1.

3. Direct and Indirect Competition

In (Glass and Schupbach 2017), we develop and defend a formal explication of hypothesis competition that satisfies both of the above desiderata. In this section, we motivate the use of this measure explicitly in terms of direct and indirect pathways to competition and we reiterate some important implications of competition explicated in this way.

Our explication is Bayesian only in the sense that it is probabilistic, and uses the relevant probabilities as formal representations of rational degrees of belief (or credences) for a particular epistemic agent. Since these are probabilities, they are assumed to satisfy the standard Kolmogorov axioms; rational agent credences are thus at least synchronically coherent in our framework. Setting up our minimal Bayesian framework, if A is an algebra of propositions (including the tautology and closed under negation and disjunction), then probability function Pr: A→[0,1] assigns the degrees of belief that an ideal Bayesian agent has in any and all members of A. The degree to which one proposition (incrementally) *confirms* another is measured differently by distinct candidate measures. Our favored measure is the "log-likelihood" measure (and other measures ordinally equivalent to it). For any three propositions x, y, z ∈ A, this measure calculates the degree to which x confirms y, given z, as follows:

$$C_l(x,y|z) = \log\left[\frac{\Pr(x|y\&z)}{\Pr(x|\neg y\&z)}\right]$$

The degree to which proposition x ***dis****confirms* y, given z, on the other hand is defined as the degree to which x confirms ¬y, given z:

$$C_l(x,\neg y|z) = \log\left[\frac{\Pr(x|\neg y\&z)}{\Pr(x|y\&z)}\right] = -C_l(x,y|z)$$

Using C_l, our measure of the overall (or "net") degree to which H and H′ compete with respect to E may be represented as the average degree to which H and H′ disconfirm each other in light of E:

$$\text{Comp}(H,H'/E) = [C_l(H,\neg H'|E)+C_l(H',\neg H|E)]/2$$

Comp takes increasingly positive values to the extent that H and H′ compete (or disconfirm one another in light of E), decreasingly negative values to the extent that H and H′ support (or confirm) one another, and value 0 exactly when H and H′ are irrelevant to one another in light of E (this happens, e.g., when the direct and indirect competition or support between H and H′ with respect to E exactly cancel each other out).[3]

We may demonstrate that Comp captures both pathways to competition by showing that it is a simple combination of separate measures of degree of direct and indirect competition respectively. Recall that competition between hypotheses H and H′ (where H, H′ ∈ A) occurs along the *direct* pathway to the extent that these hypotheses confirm the falsity of each other (i.e., "disconfirm" each other). In our previous paper, we used degree of disconfirmation $C_l(H',\neg H)$ to represent the degree to which H′ directly rebuts H and used this as part of our account of competition. Here we extend this by using C_l to formalize degree of direct competition Comp_D as the average degree to which either hypothesis confirms the negation of the other:

$$\text{Comp}_D(H,H') = [C_l(H',\neg H)+C_l(H,\neg H')]/2 = \left(\log\left[\frac{Pr(H'|\neg H)}{Pr(H'|H)}\right] + \log\left[\frac{Pr(H|\neg H')}{Pr(H|H')}\right]\right)/2$$

Next, recall that competition between H and H′ occurs along the *indirect* path, by way of some body of evidence E ∈ A, to the extent that these hypotheses undermine the support that E provides for each other. To measure this extent, one must compare the degree to which E confirms either hypothesis (without consideration for the other hypothesis) with the degree to which E confirms that same hypothesis after the alternative hypothesis is accepted or assumed. For example, E confirms H to degree $C_l(E,H)$; given H′, however, E confirms H to degree $C_l(E,H|H')$. In our previous paper, we defined the extent to which H′ turns E against H by subtracting the latter from the former: $C_l(E,H)-C_l(E,H|H')$. Here we use this

[3] Comp lies in the range [-∞,+∞]. For a measure of competition with range [−1,1], one can instead opt for measure $[C_k(H,\neg H'|E)+C_k(H',\neg H|E)]/2$, which uses the ordinally equivalent "Kemeny-Oppenheim" measure of incremental confirmation:
$C_k(x,y|z) = [\Pr(x|y\&z)-\Pr(x|\neg y\&z)]/[\Pr(x|y\&z)+\Pr(x|\neg y\&z)]$.

approach to define degree of indirect competition $\mathrm{Comp_I}$ as the average degree to which the hypotheses turn E against one another:

$$\mathrm{Comp_I}(H,H'/E) = [C_I(E,H)-C_I(E,H|H')+C_I(E,H')-C_I(E,H'|H)]/2$$
$$= \left(log\left[\frac{Pr(E|H)}{Pr(E|\neg H)}\right] - log\left[\frac{Pr(E|H\&H')}{Pr(E|\neg H\&H')}\right] + log\left[\frac{Pr(E|H')}{Pr(E|\neg H')}\right] - log\left[\frac{Pr(E|H'\&H)}{Pr(E|\neg H'\&H)}\right]\right)/2$$

The desired result—that Comp captures both pathways to competition—then follows, as expressed in the following theorem:

Theorem 1. $\mathrm{Comp}(H,H'/E) = \mathrm{Comp_D}(H,H') + \mathrm{Comp_I}(H,H'/E)$.[4]

Using our measure, we can make more precise sense of qualitative judgments of hypothesis competition. We explicate the judgment that H and H′ compete with respect to E (to some degree) as a positive degree of competition, $\mathrm{Comp}(H,H'/E)>0$. Since positive degree of competition corresponds to the case in which H and H′ disconfirm one another in light of E (and ignoring cases in which Pr(H) or Pr(H′) is zero), we may also state the (qualitative) condition for competition using the following probabilistic inequality

$$Pr(H|H'\&E)<Pr(H|E) \tag{1}$$

In an earlier study on "explaining away," Glass (2012, theorem 1) also proved that the following inequality provides another equivalent condition for competition:

$$\log\left[\frac{Pr(E\,|\,H\&H')\times Pr(E|\neg H\&\neg H')}{Pr(E\,|\,H\&\neg H')\times Pr(E|\neg H\&H')}\right] + \log\left[\frac{Pr(H|H')\times Pr(\neg H|\neg H')}{Pr(H|\neg H')\times Pr(\neg H|H')}\right] < 0 \tag{2}$$

We can reapply (2) for our own purposes here, giving us another statement of the criterion for competition (to some degree). (2) turns out to be especially useful insofar as it, unlike the

---

[4] Proof. [This follows from theorem 1 in Glass and Schupbach (2017).]

$\mathrm{Comp_D}(H,H') + \mathrm{Comp_I}(H,H'/E)$

$$= \left(\log\left[\frac{Pr(H'|\neg H)}{Pr(H'|H)}\frac{Pr(E|H)}{Pr(E|\neg H)}\frac{Pr(E|\neg H\&H')}{Pr(E|H\&H')}\right] + \log\left[\frac{Pr(H|\neg H')}{Pr(H|H')}\frac{Pr(E|H')}{Pr(E|\neg H')}\frac{Pr(E|\neg H'\&H)}{Pr(E|H'\&H)}\right]\right)/2$$

$$= \left(\log\left[\frac{Pr(H'|\neg H)\,Pr(E|\neg H\&H')}{Pr(E|\neg H)}\frac{Pr(E|H)}{Pr(H'|H)\,Pr(E|H\&H')}\right] + \log\left[\frac{Pr(H|\neg H')\,Pr(E|\neg H'\&H)}{Pr(E|\neg H')}\frac{Pr(E|H')}{Pr(H|H')\,Pr(E|H'\&H)}\right]\right)/2$$

$$= \left(\log\left[\frac{Pr(H'|\neg H\&E)}{Pr(H'|H\&E)}\right] + \log\left[\frac{Pr(H|\neg H'\&E)}{Pr(H|H'\&E)}\right]\right)/2 = [C_I(H',\neg H|E)+C_I(H,\neg H'|E)]/2 = \mathrm{Comp}(H,H'/E).$$

other qualitative criteria, separates *independent* terms (the two summands) relating to the indirect and direct pathways by which competition can occur respectively.[5]

Figure 2 displays degree of competition as a function of the likelihoods for the hypotheses on their own—i.e. Pr(E|¬H&H′) and Pr(E|H&¬H′)—in a scenario where the hypotheses are independent before the evidence is taken into account, corresponding to the absence of direct competition as displayed in figure 1(a), and to the case where the second term in (2) is zero. The values of the probabilities Pr(H|E) and Pr(H|H′&E) are also shown in the figure. For the limiting case in which the likelihoods are zero, the first term in (2) is infinite and so the degree of competition is minimal. Essentially, this means that neither hypothesis can account for the evidence on its own, since either hypothesis alone (i.e., conjoined with the negation of the other hypothesis) implies the falsity of E. Furthermore, when the likelihoods are zero, Pr(H|H′&E) = 1 and so given E, H′ guarantees the truth of H. As the likelihoods increase, so does the degree of competition until it reaches zero when Pr(H|H′&E) = Pr(H|E), which occurs at a value of the likelihood close to 0.25. This corresponds to the first term in (2) being zero. For greater values of the likelihoods, Pr(H|H′&E) < Pr(H|E) and so the degree of competition is positive. This corresponds to competition along pathway 1.

Figure 3 shows another scenario similar to the previous one, but where there is a direct dependence between the hypotheses corresponding to figure 1(b). Negative dependence between the hypotheses, which occurs for Pr(H′|H) < 0.1, gives rise to competition via pathway 2 and so adds to the competition due to pathway 1. For sufficiently large positive dependence the negative contribution via pathway 1 is cancelled out via the contribution from pathway 2 resulting in no competition. This occurs at Pr(H′|H) = 0.3 when Pr(H|H′&E) = Pr(H|E) and as the dependence between H′ and H increases further the degree of competition becomes lower still.

[5] That is, if a probabilistic model is specified in terms of the conditional probabilities appearing in (2), then the two summands in (2) can be varied independently.

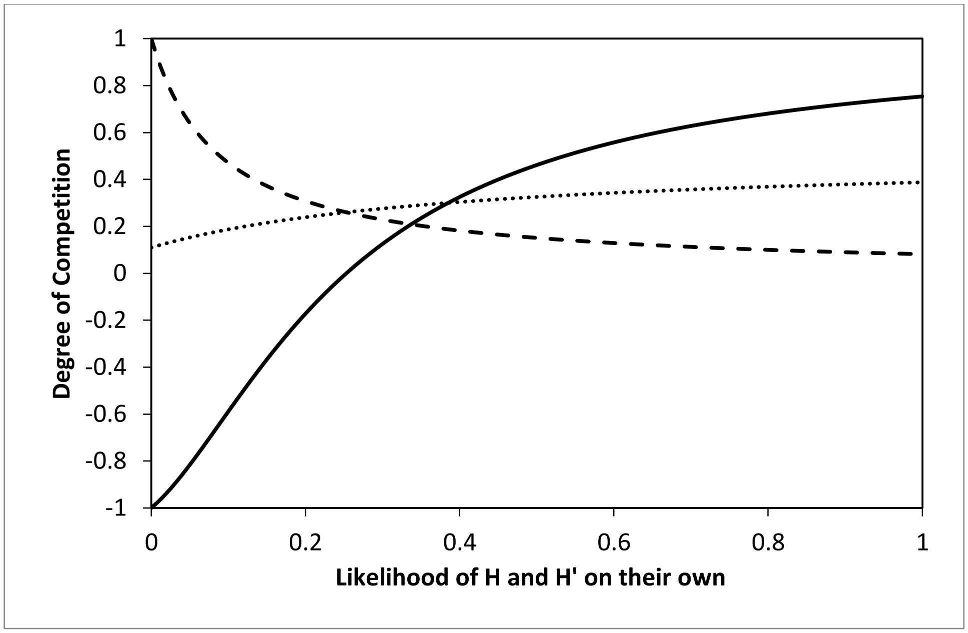

Figure 2. Degree of competition (solid line) as a function of the likelihoods Pr(E|H&¬H′) = Pr(E|¬H&H′), when Pr(H) = Pr(H′) = 0.1, Pr(E|H&H′)= 0.8 and Pr(E|¬H&¬H′) = 0.08. Also shown are Pr(H|H′&E) (dashed line) and Pr(H|E) (dotted line). Note that here we use an ordinally equivalent measure of competition with range [-1,1] (constructed using C_k as described in footnote 3).

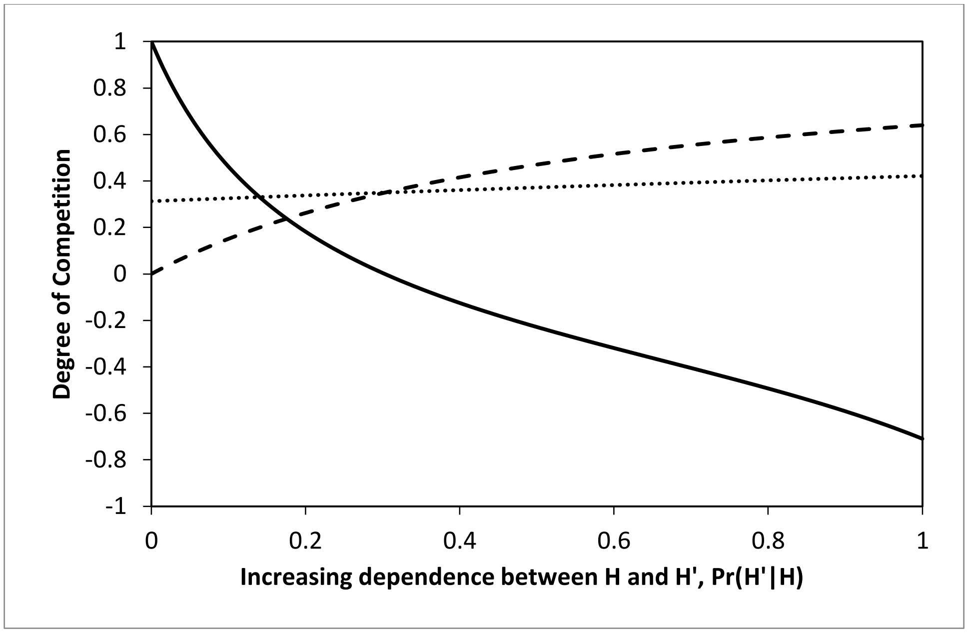

Figure 3. Degree of competition (solid line) as a function of dependence between hypotheses. Pr(H′|H) varies from 0 to 1 while Pr(H′|¬H) is fixed at 0.1; thus, dependence is negative for Pr(H′|H) < 0.1 and positive for Pr(H′|H)> 0.1. Apart from Pr(H′) other probabilities are as defined in figure 2 except that now Pr(E|H&¬H′) and Pr(E|¬H&H′) are fixed at 0.5. Also shown are Pr(H|H′&E) (dashed line) and Pr(H|E) (dotted line). Note that here we use an ordinally equivalent measure of competition with range [-1,1] (constructed using C_k as described in footnote 3).

4. Computer Simulations of Bayesian inference

In Section 2, we argued that the mutual exclusivity assumption, though common in formal philosophy of science research, is often not true to scientific practice (the object of philosophy of science research). At best then, this assumption is a relatively innocuous idealizing assumption. At worst, it perniciously confounds a wide swath of work in contemporary philosophy of science. In this section, we present some initial research into the important questions of just how harmful this idealizing assumption can be to formal results, and under what conditions.

In order to investigate the impact of incorrectly assuming that hypotheses are mutually exclusive on making inferences, we carried out computer simulations. The idea is first to define a probability model, which we will stipulate as the correct model, Pr, involving H, H′ and E, where H and H′ are not assumed to be mutually exclusive. We then modify this model to obtain what we will call the mutually exclusive probability model, Pr_X, where H and H′ are treated as mutually exclusive. The next step is to make inferences using the mutually exclusive model and determine how good they are by comparing them with inferences made using the correct model. Inferences are made by selecting the most probable hypothesis given E, so essentially we evaluate the mutually exclusive model by determining how often it identifies the same hypothesis as the correct model—or how often it correctly identifies the most probable of potentially competing hypotheses. Here we focus on how the results depend on the direct competition between the two hypotheses and so the above steps are repeated for different values of direct competition.

The first step is to define the correct probability model, Pr, for a given degree of direct competition between the hypotheses. Since the measure of direct competition, Comp_D, lies in the range [-∞,∞], to obtain a measure with range $[-1,1]$, we instead use the measure $[C_k(H,\neg H'|E)+C_k(H',\neg H|E)]/2$, which uses the ordinally equivalent "Kemeny-Oppenheim" measure of incremental confirmation, $C_k(H',\neg H) = [\Pr(H'|\neg H) - \Pr(H'|H)]/ [\Pr(H'|\neg H) + \Pr(H'|H)]$. For a given value of direct competition, d, we can express d as $(d_1 + d_2)/2$, where $d_1 = (b_1 - a_1)/(b_1 + a_1)$, where $a_1 = \Pr(H|H')$ and $b_1 = \Pr(H|\neg H')$ and $d_2 = (b_2 - a_2)/(b_2 + a_2)$, where $a_2 = \Pr(H'|H)$ and $b_2 = P(H'|\neg H)$. For a positive value of d, (a similar approach is adopted for negative values), d_1 can then be selected randomly from a uniform distribution over the interval (max(2d - 1),0), min(2d,1)) and d_2 then set to $2d - d_1$. Pr(H) is selected randomly. With Pr(H) fixed, we can obtain values for Pr(H′), a_1, b_1, a_2 and b_2—since (i) $\Pr(H) = a_1\Pr(H') + b_1\Pr(\neg H')$, where $b_1 = a_1(1 + d_1)/(1 - d_1)$, (ii) there are corresponding expressions for Pr(H′) and b_2, and (iii) $a_1\Pr(H') = a_2\Pr(H)$.[6]

As was the case for figure 3, we select values of likelihoods so that $\Pr(E|\neg H\&\neg H') < \min[\Pr(E|H\&\neg H'), \Pr(E|\neg H\&H')] \leq \max[\Pr(E|H\&\neg H'), \Pr(E|\neg H\&H')] < \Pr(E|H\&H')$. Of course, other assignments of likelihoods are possible, but this reflects the situation where the evidence is more likely if either one of the hypothesis is true rather than if neither is true and it is more likely still if both hypotheses are true. It also ensures that both hypotheses are confirmed by E, i.e. $\Pr(H|E) > \Pr(H)$ and $\Pr(H'|E) > \Pr(H')$ in the case where H and H′ are independent.

Now that the correct probability model has been specified, it needs to be modified to obtain the mutually exclusive probability model, $\Pr_X$. We define prior probabilities for H and H′ so that $\Pr_X(H \vee H') = \Pr_X(H) + \Pr_X(H') = \Pr(H \vee H')$, which means that the total area of the probability space taken up by the hypotheses remains the same. This is achieved by reducing each correct prior $\Pr(H_i)$ by multiplying it by the same factor $\Pr(H \vee H') / (\Pr(H)+\Pr(H'))$. Apart from that, we simply set $\Pr_X(E|H) = \Pr_X(E|H\&\neg H') = \Pr(E|H\&\neg H')$, $\Pr_X(E|H') = \Pr_X(E|\neg H\&H') = \Pr(E|\neg H\&H')$ and $\Pr_X(E|\neg H\&\neg H') = \Pr(E|\neg H\&\neg H')$.

---

[6] There are constraints on how Pr(H) is selected to guarantee a coherent probability distribution, but this can be implemented simply by selecting Pr(H) from the interval (0,1) until the distribution is coherent. The value of Pr(H) does not suffice to determine the value of Pr(H′) and hence a_2 and b_2 in cases where H and H′ are independent and so $d = d_1 = d_2 = 0$. In this case the value of Pr(H′) is also selected randomly from (0,1).

With the two models now in place we find the hypothesis that maximizes $Pr_X(H_i|E)$. We compare it with the hypothesis that maximizes $Pr(H_i|E)$ and, if they match, count the iteration as a success for the mutually exclusive model. By repeating this process (10^7 times in the computer simulations that were carried out) of selecting the models and maximizing the posterior probability, we are able to determine the accuracy of the mutually exclusive approach as the percentage of cases where it is successful. As noted earlier, this process is then repeated for different values of direct competition between H and H′.

Clearly, when the mutually exclusive approach is used, only one inference can be made, but in reality (as described here by the correct model) both hypotheses could be true unless the degree of competition is maximal. To get an idea of how much of a weakness this is in the mutually exclusive approach, we also obtained the accuracy by considering the number of successes for hypotheses whose posterior probability in the correct model is greater than 0.5 as a percentage of the total number of such hypotheses. This can be expressed as

$$\frac{\text{Total number of correctly identified hypotheses with probability greater than 0.5}}{\text{Total number of hypotheses with probability greater than 0.5}} \times 100\%$$

where the probability refers to the posterior distribution $Pr(\cdot|E)$ of the correct model.

Results are presented in figure 4. Consider first of all the results for the general case (solid line). It is clear that as the degree of direct competition approaches its maximal value of 1 the accuracy of making inferences using the mutually exclusive model approaches 100%. This makes sense since maximal competition corresponds to mutual exclusion. While the accuracy of the mutually exclusive approach is greater than 97% for degrees of competition above about 0.5, in general the accuracy decreases as degree of direct competition decreases and is at 70% for values close to -1, which corresponds to the case where the hypotheses, far from competing, entail each other. While this seems like a reasonable level of accuracy, it must be remembered that there are only two hypotheses to choose from and so a random guess would achieve an accuracy of 50%. Note that a degree of competition of 0.5 corresponds to cases in which each hypothesis is on average still three times as likely to be true if the alternative hypothesis is false, $Pr(H_i|\neg H_j) = 3 \times Pr(H_i|H_j)$ while a degree of direct competition of 1/3 corresponds to the case where each is on average still twice as likely. In such cases, there remains a substantial relation of negative relevance between the hypotheses, and thus reasoners may still be strongly inclined to think of them as competitors.[7]

[7] Note that for both sets of results in figure 4 the results for a zero degree of competition are higher than the general trend might suggest, which is presumably related to the fact that there is an independence relationship present in this case that is absent in the other cases. Because of this independence the probability distribution had to be selected in a different way in this case as mentioned in footnote 6.

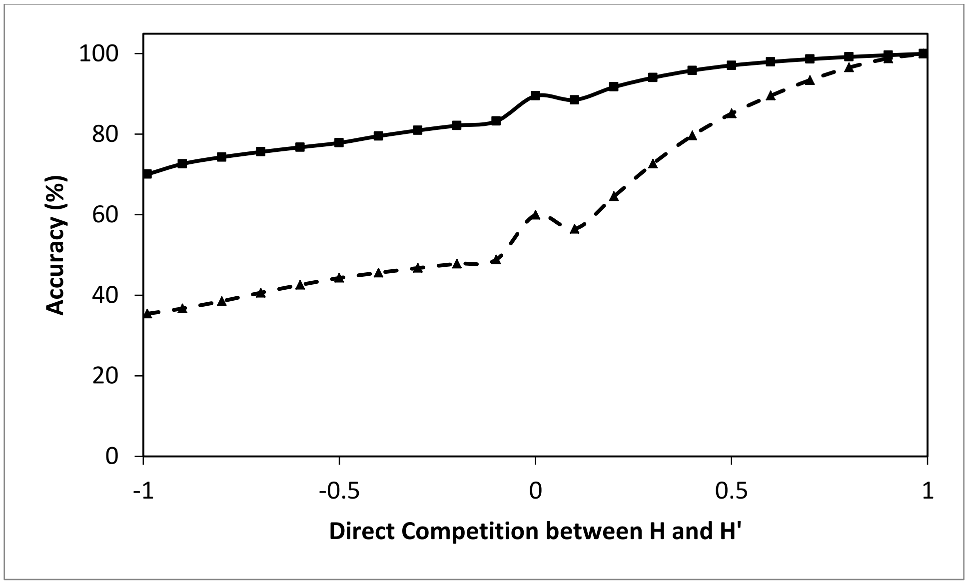

Figure 4. Accuracy of inferences made using the mutually exclusive model as a function of the degree of direct competition between the hypotheses. Results are presented for the general case (solid line) and for the case where all hypotheses with a posterior probability greater than 0.5 are taken into account in determining accuracy (dashed line).

Now consider the results that take into account cases where both hypotheses H and H′ have a posterior probability greater than 0.5 according to the correct model (dashed line in figure 4). When the degree of competition is 1 the hypotheses cannot both have probability greater than 0.5, i.e. they cannot both be more likely to be true than false. Hence for high degrees of competition, the accuracy is still close to 100%. However, the accuracy now falls off much more quickly so that it is already below 75% when the degree of competition is 0.3. In general, the accuracy continues to fall with lower values of competition until it reaches an accuracy of just 35% as the degree of competition approaches -1. Of course, it is not surprising that the mutually exclusive approach performs worse in this case since according to it at most one hypothesis could be true (or have probability greater than 0.5). These results highlight the extent of perhaps the most significant problem with the mutually exclusive approach: its exclusion of hypotheses that may well be true.

In summary, insofar as inference is concerned with identifying the most probable hypothesis, it could be argued from the results in figure 4 that the mutually exclusive approach performs quite well, with an accuracy close to or greater than 90%, provided the hypotheses are in direct competition to at least some extent, i.e. the degree of direct competition is greater than zero. However, the problem of excluding hypotheses that may well be true is a problem even in cases where the hypotheses are competing. Of course, all of these results provide insight into cases where we are only monitoring varying degrees of *direct* competition; future investigations will explore the extent to which these inaccuracies might be compounded when we instead consider cases of varying degrees of *indirect* and/or *net* competition.

5. Case Study

Let us return to our earlier discussion of consistent but competing scientific hypotheses in the context of the mass extinction at the K-Pg boundary. Here we focus on two of the causal hypotheses, asteroid impact and Deccan volcanism, which we denote as H and H′ respectively. There is no reason to think that these hypotheses are mutually exclusive. While there is no competition along the direct pathway 2, there may be competition along the indirect pathway 1. If so, the discovery of evidence that provides confirmation of the impact hypothesis, as noted earlier, would count against volcanism via pathway 1 even if this evidence does not directly contradict volcanism. This is the main strategy adopted by Schulte et al. (2010) who, assuming competition between the hypotheses, argue against volcanism essentially on the grounds that it is not needed because the asteroid hypothesis on its own accounts for the relevant evidence.

Some responses to Schulte et al. suggest that the impact and volcanism hypotheses are not in competition (Archibald et al. 2010; Courtillot and Fluteau 2010; Keller et al. 2010). Their arguments, as well as other work (Brusatte 2015), which relate to the first term in expression (2) and hence to pathway 1, amount to the claim that overall the evidence given both impact and volcanism hypotheses (and perhaps other hypotheses too) is much more probable than it is given the impact hypothesis without volcanism, i.e. $Pr(E|H\&H')$ is much greater than $Pr(E|H\&\neg H')$. In a recent paper, Renne et al. (2015) argue that the Chicxulub impact accelerated the rate of volcanic activity from the Deccan Traps. This kind of direct positive dependence between the hypotheses would result in a positive second term in (2) and would lead to support between the hypotheses rather than competition along pathway 2.

This very brief survey indicates that there are grounds for questioning the assumption that the impact and volcanic hypotheses are competing. Furthermore, even if the hypotheses are competing to some degree, the simulation results presented in section 4 show that if the assumption of mutual exclusion is made for inference purposes it could easily lead to exclusion of a true hypothesis.

6. Conclusion

We have argued that hypothesis competition cannot be adequately understood in terms of mutual exclusion. It is important to recognize that competition is a matter of degree and that it can occur in different ways – directly or indirectly via the evidence. In Section 3, we defended a previously proposed measure by showing how it explicates these features in terms of degrees of direct and indirect competition and by illustrating its quantitative behavior in simple scenarios. It might be thought that mutual exclusivity would suffice as a simplifying assumption that would capture key aspects of scientific reasoning, but the results of computer simulations in Section 4 show that this approach can have significant limitations, particularly resulting in the exclusion of hypotheses that may well be true. Finally, our brief discussion of the extinction of the dinosaurs in Section 5 illustrates subtleties of competition in scientific practice that go well beyond mutual exclusion. This work suggests that there is a need for philosophers of science to explore the nature of hypothesis competition in more detail and we hope that our account provides a helpful framework for this task.

References

Alvarez, Luis W., Walter Alvarez, Frank Asaro, and Helen V. Michel. 1980. "Extraterrestrial cause for the Cretaceous-Tertiary extinction." *Science* 208:1095–1108.

Archibald, J. David, William A. Clemens, Kevin Padian et al. 2010. "Cretaceous extinctions: multiple causes." *Science* 328:973.

Benton, Michael J. 1990. "Scientific methodologies in collision: the history of the study of the extinction of the dinosaurs." *Evolutionary Biology* 24:371–400.

Brusatte, Stephen L. et al. 2015. "The extinction of the dinosaurs." *Biological Reviews* 90:628-642.

Cleland, Carol E. 2002. "Methodological and epistemic differences between historical science and experimental science." *Philosophy of Science* 69:474-496.

———. 2011. "Prediction and explanation in historical natural science." *British Journal for the Philosophy of Science* 62:551–582.

Courtillot, Vincent, and Frédéric Fluteau. 2010. "Cretaceous extinctions: the volcanic hypothesis." *Science* 328:973–974.

Glass, David H. 2012. "Can evidence for design be explained away?" In *Probability in the Philosophy of Religion*, ed. J. Chandler, and V. Harrison, 79–102. Oxford: Oxford University Press.

Glass, David H., and Jonah N. Schupbach. 2017. "When do hypotheses compete?" Under Review.

Hildebrand, Alan R., Glen T. Penfield, David A. King et al. 1991. "Chicxulub crater: a possible Cretaceous/Tertiary boundary impact crater on the Yucatan Peninsula, Mexico." *Geology* 19:867–871.

Keller, Gerta, Thierry Adatte, Alfonso Pardo et al. 2010. "Cretaceous extinctions: evidence overlooked." *Science* 328:974–975.

Renne, Paul R., Courtney J. Sprain, Mark A. Richards et al. 2015. "State shift in Deccan volcanism at the Cretaceous-Paleogene boundary, possibly induced by impact." *Science* 350:76-78.

Schulte, Peter, Laia Alegret, Ignacio Arenillas et al. 2010. “The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary.” *Science* 327:1214–1218.