INTERNATIONAL STUDIES IN THE PHILOSOPHY OF SCIENCE
https://doi.org/10.1080/02698595.2022.2084705

On the Elusive Formalisation of the Risky Condition for Hypothesis Testing

José Díez and Albert Solé

BIAP, LOGOS Research Group, University of Barcelona

ABSTRACT
In this paper, we examine possible formalisations of the riskiness condition for hypothesis testing. First, we informally introduce derivability and riskiness as testing conditions together with the corresponding arguments for refutation and confirmation. Then, we distinguish two different senses of confirmation and focus our discussion on one of them with the aid of a historical example. In the remaining sections, we offer a brief overview of the main references to the risky condition in the literature and scrutinise different options for formally capturing riskiness; we show why none of them works. We conclude with some remarks about the relation between derivability and riskiness and claim that riskiness essentially involves a contextual, pragmatic component that eludes a complete formal reconstruction.

Many scientific hypotheses make claims that go beyond the observable evidence. How can these be assessed? As is well known, hypotheses are tested indirectly by means of their observable consequences. Intuitively, a good test is such that if a predicted consequence is shown to obtain, then we can be confident about the hypothesis; whereas if the consequence fails, then we must question the hypothesis.

For this procedure to work, two conditions must be met. If H is the non-directly observable hypothesis, P is the (taken as) observable phenomenon,[1] A are the auxiliaries and IC the initial experimental conditions,[2] then the two conditions required for hypothesis testing may be informally formulated as follows:

(DER) Derivability of the evidence: H and A and IC together imply P.

(RISK) Riskiness/unlikelihood of the evidence: If H were false, but A and IC true, then it is very likely that P would not obtain; that is, the prediction must be 'risky': not easily expectable without the hypothesis.

The intuitive confirmatory, ampliative argument that inductively supports the hypothesis in case of the success of a risky prediction has this form:

If not-H & A & IC then very likely not-P
P
A & IC

Q1 **CONTACT** José Díez diez.ja@gmail.com BIAP, LOGOS Research Group, University of Barcelona

THEN INDUCTIVELY:
H

Most philosophers agree that this argument is intuitively satisfactory. Nevertheless, it is not obvious how it must be read precisely, for the schema 'If X then very probably Y; not-Y; THEREFORE, INDUCTIVELY not-X' is *not* in general a good inductive argument. The question then arises as to why or under what conditions the confirmatory argument above is inductively valid. Giere acknowledges that the inductive version of *modus tollens* is not in general valid, but claims that if derivability also obtains, then the confirmatory argument is indeed valid (Giere 1979, 96–99). As we will show (Section 4), this proposal is unsatisfactory, and thus the question remains of how one has to understand this argument, and more particularly the risky condition that constitutes its first premise, for it to be valid, as intuitively it is. This is the issue we explore in the present paper.

In Section 1, we introduce the two conditions outlined above in more detail, as well as the arguments on which the refutation and the confirmation of a hypothesis are based. It is easy to see that the notion of confirmation that underlies the risky condition and the confirmatory argument above does not allow for two incompatible hypotheses to be confirmed by the same evidence. In Section 2, we maintain that it is this sense of confirmation that is at stake in several paradigmatic cases of hypothesis testing. We then discuss one historical episode which motivates this claim and in order to illustrate the main features of the risky condition involved. An important part of this condition—or so we argue—is the requirement that no other *viable* rival hypothesis make the same prediction in the given context. In Section 3, we offer an overview of some of the main references to the risky condition in the literature, noting that, despite its importance, this latter requirement is rarely made explicit. Through Sections 4–7, we explore several attempts to provide a formal elucidation of this condition with negative results. More particularly, in Section 4, we argue that the content of the risky condition cannot be properly expressed by resorting to material conditionals. In Section 5, we proffer reasons for being sceptical about satisfactorily explicating riskiness in terms of conditional probabilities. Then, in Sections 6 and 7, we scrutinise formalisations of riskiness that involve probabilistic thresholds or explicitly refer to alternative hypotheses and their probabilities. Finally, in Section 8, we sum up our results and conclude that riskiness involves an ineliminable contextual and pragmatic component that eludes a complete formal reconstruction; we also conclude that, as it is used in confirmatory arguments, the risky condition should be read as including a positive probabilistic relation between the hypothesis and the prediction.

1. Derivability and Riskiness, Refutation and Confirmation

As is well-known, (DER) is used to refute H when both A and IC hold but P does not in fact occur, according to the following valid deductive argument:

(REF)

$$H \wedge A \wedge IC \rightarrow P$$
$$\neg P$$
$$A \wedge IC$$
$$\overline{\hspace{6em}}$$
$$\neg \mathrm{H}$$

While (DER) is sufficient to refute H if the prediction P fails and A and IC are true, it is obviously not enough to confirm H if the predicted P obtains. If the predicted phenomenon, P, is something to be expected regardless of the occurrence of H (perhaps because P is highly probable in itself, or because there are other plausible hypotheses in the context that also predict P), then it is incorrect to take H as confirmed by the occurrence of P. For this reason, as is well known, confirmation requires the additional condition (RISK), which is meant precisely to exclude those scenarios. As we have seen, with (RISK) at hand, one can construct a confirmatory ampliative argument which takes (RISK), P, A and IC as premises, and inductively concludes H, with a degree of inductive force corresponding to the likelihood affirmed in the first premise. Let us write that argument out again (we use a discontinuous line to denote partial support in inductive inferences, to be duly contrasted with the total support of deductive inferences denoted by a continuous line, as above):

(CONF)

If $\neg H \wedge A \wedge IC$ then-very-likely $\neg P$
P
$A \wedge IC$

H

This schema constitutes the intuitively sound, ampliative inference in virtue of which H is confirmed if P obtains and A and IC do not fail. Of course, the inference is sound *provided* that one assumes the correct reading of the first premise (RISK). Part of this paper is devoted precisely to unravelling the correct reading of this riskiness condition (the hyphenation of 'then-very-likely' in (CONF) is intended to connote that this is a primitive expression that has yet to be analysed). As is well known, given its ampliative nature, even in the strongest inductive confirmations the premises may all be true but the hypothesis be false, as the history of science has often dramatically demonstrated.

The interplay between conditions (DER) and (RISK), and arguments (REF) and (CONF) is somehow intriguing. On the one hand, (RISK) is not involved in (REF) and, in fact, the argument (REF) is also cogent even when (RISK) fails: if H implies P and P fails, H is questioned regardless of how probable P is or how many alternative hypotheses also predict P. So, clearly, riskiness is not necessary for refutation. On the other, more intriguing, hand, since (DER) does not explicitly appear among the premises of (CONF), it would *seem* that, when (RISK) holds, we are in a position to confirm H even if (DER) does not hold. However, we have mentioned that Giere (1979, 99) assumes that the confirmatory argument is inductively valid if (DER) holds. We will show the inadequacy of such a claim, yet it is still true that it would not make sense for a hypothesis, H, to be confirmed by a piece of evidence, P, that is not *appropriately related* to H. So, something like (DER)—though not necessarily (DER) itself—seems intuitively necessary also for confirmation. Whether this should be added as a premise of (CONF) or not, depends on what the precise interpretation of (RISK) is, for it might happen that something like (DER) is already an implicit part of the content of (RISK). We will come back to this point later on, when we scrutinise the precise content, and possible formalisation, of (RISK).

It is furthermore well known that the form of (DER) is that of a logical implication: P deductively follows from $H \wedge A \wedge IC$ (this is why the auxiliaries are characterised as what one has additionally to assume for H to imply P in conditions IC). But notice that we have written (REF) as is usual: with the first premise *not* being exactly (DER), but the weaker material implication '$H \wedge A \wedge IC \rightarrow P$' to which *modus tollens* applies. Of course, if (DER) is true, then $H \wedge A \wedge IC \rightarrow P$ also is; so (DER) ensures that *modus tollens* can be applied. But given that the weaker material implication is enough to refute H when P does not obtain (but A and IC do), one may wonder why the testing of H requires the stronger (DER). The reason why this must be so is rarely made explicit in the literature. For us, the key has to do with *epistemic accessibility*. When one proceeds to hypothesis testing, one should know *before* checking P, that P is indeed relevant for assessing H. That is, one should know in advance of performing the test that the two conditions that allow for refutation-in-case-of-failure and for confirmation-in-case-of-success, obtain. But if (DER) were a material conditional, this could not be the case. A material conditional is true when the consequent is true or the antecedent false. Therefore, in order to know that $H \wedge A \wedge IC \rightarrow P$ is true, we should know either that P is true or that $H \wedge A \wedge IC$ is false. But one cannot know this disjunction before the test.[3] This epistemic status dramatically changes when we move from material to logical implication, since the latter is knowable *a priori*, thus epistemically accessible before performing any test. One can then know perfectly well whether (DER) is true before knowing if P obtains, since (DER) claims that P deductively follows from the conjunction of H and A and IC, which in principle can be known by applying the rules of logic and/or performing mathematical calculation. In fact, the only general way of knowing that $H \wedge A \wedge IC \rightarrow P$ is true before knowing whether P is true, is just by knowing that (DER) holds.

This epistemic constraint applies to both conditions, so (RISK) must also be epistemically accessible before performing the test. And actually, this is so; for the fact that P is unlikely or that it is not predicted by other hypotheses is assessed on the basis of background information already available before testing H. This does not mean, though, that in this case too merely logical/mathematical information suffices to know whether (RISK) is satisfied on a given occasion. The background knowledge required to assess the plausibility of P in scenarios where H does not hold is not *a priori* and, as we will argue, in this case some contextual and pragmatic information is essential and suggests that riskiness eludes a complete formal explication.

Another consideration worth mentioning at this point concerns the notion of confirmation that is implied by an argument like (CONF), which is a valid inductive inference that has P among its premises and the hypothesis H as its conclusion. Therefore, the evidence does not merely increment the probability of the hypothesis, but the former renders the latter highly probable and thus worth believing/accepting after the test. According to this notion of confirmation, let us call it 'Conf1', the same evidence cannot confirm two or more contradictory hypotheses that predict it. It would be clearly nonsensical to accept that in virtue of one and the same test we are justified in believing that two (or more) contradictory hypotheses are very likely to be true on the basis of (CONF), since in such a situation the first premise (correctly appraised) simply cannot be true of any of these hypotheses.

We take it that Conf1 is precisely the notion that Earman and Salmon (1992, 89) refer to as a 'high probability' or 'absolute' notion of confirmation, and it is the one that we will be focusing on in what follows. There is another notion of confirmation, let us call it 'Conf2', according to which 'positive' evidence for a given hypothesis (in a sense to be specified) does confirm the hypothesis to a certain degree, regardless of the presence of other hypotheses that entail the same prediction. In this sense, the same piece of evidence may 'confirm' different, even contradictory, hypotheses at the same time. This notion of confirmation has been referred to as 'incremental confirmation' by Earman and Salmon (1992, 89) and it is the one at stake when, for instance, in Bayesian style, that evidence P confirms H whenever Pr(H|P) is greater than Pr(H).

Given this plurality of concepts, one could object that it is Conf2—and not Conf1—that is relevant to understanding classic episodes of hypothesis testing in the history of science. In the next section, we argue that this is not the case, and that it is precisely Conf1 that is at work in paradigmatic historical cases of hypotheses testing. In order to do so, we present the historical episode of the confirmation of the Copernican system by Galileo's observation of the phases of Venus in a certain degree of detail. We also use this historical case to help clarify exactly what is involved in the risky condition, which is the object of our analysis throughout the rest of the paper.

2. Riskiness and Conf1: Galileo and the Phases of Venus

We maintain that the two notions of confirmation referred to above, Conf1 and Conf2, are neither in conflict nor in competition with each other, but simply serve different purposes. Conf2 is the concept at work, for instance, the Bayesian updating machinery, which accounts *after the fact* for the problem of the possible presence of alternative hypotheses making the same prediction via convergence theorems (see e.g. Sprenger 2011 for what we take to be a Conf2 approach). We do not question this notion or its corresponding analysis. Yet, as we have advanced, we consider that Conf1 is the relevant concept in many 'belief formation' or 'action grounding' contexts, where one must decide what to believe, and how to act accordingly. We take it that, typically, Conf1 is the notion of confirmation at issue when scientists claim that a theory or hypothesis has been confirmed and therefore (provisionally) established as correct after successfully passing a test; and that it is also used by applied scientists, technicians, physicians, attorneys, detectives, and the like.

One may object to our Conf1/Conf2 distinction and claim that Conf2 is also applicable to our intended cases, for, if there are several hypotheses making the same successful prediction, then each hypothesis-community simply takes its own hypothesis as the one to have been confirmed.[4] We believe, however, that this may correspond to two different scenarios. In the first scenario, each sub-community accepts (for whatever reasons) as 'worthy to be assessed' only its own respective hypothesis and does not even entertain the hypotheses considered by the other communities. In this scenario, of course, each of these communities will take its own hypothesis to be confirmed after it has successfully passed the test. Notice, however, that this is the case precisely because the rival hypotheses are not even considered worthy of being taken into account and therefore the risky condition (RISK) holds in the context of each community, as applied to its own hypothesis. Therefore, each community applies (CONF) and considers its hypothesis to have been confirmed. In this scenario, though, the notion at stake in each community is clearly Conf1, *not* Conf2. In a second scenario, within one and the same community, two or more hypotheses are taken to be worth considering before a test. Imagine, for instance, a practical context in which technicians must take a certain course of action and want to know which of two hypotheses, H1 or H2, that imply incompatible courses of action, they should accept; then, a surprising fact is observed which turns out to be predicted by both hypotheses! In such a situation, we contend that the observed fact is simply of no use at all with regard to the *choice* of what to do, what to believe or how to act (provided that the disjunctive belief/action is not an alternative), and that the technicians would (independently of their prior preferences) not choose one option over the other, or at least not based on the common prediction. Here, again, Conf1 is the sense of confirmation in play.

Thus, we contend that in contexts in which one must take a theoretical or practical decision between incompatible alternatives based on a test, it is Conf1 that matters. And we think that these situations are frequent in scientific contexts. We will now introduce, as a paradigmatic example, the historical episode of the observation of the phases of Venus, considered by Galileo and many others as confirmation of heliocentrism. We argue that in this episode it is Conf1 that is clearly at stake. We do not wish to claim that this is the case in all historical episodes; it suffices for us here that this case is not an odd exception but paradigmatic of a class of relevant historical episodes. If this is so, (CONF) needs to be taken into account, thus making its crucial first premise (RISK) worth analysing.

In March 1610, Galileo published *Siderus Nuncius*, in which he summarises the main astronomical discoveries made by him through pointing the recently invented telescope to the skies. He used many of these observations (sunspots, lunar valleys, Jupiter's moons, and others) to refute core assumptions of the Ptolemaic system used by Aristotelians to question Copernicanism. For instance, Aristotelians famously considered that the supralunar world is 'perfectly uniform'; a claim obviously challenged by the observations of sunspots and the irregular orography of the Moon. Aristotelians also assumed that a planet cannot have another celestial body revolving around it; something clearly refuted by the discovery of Jupiter's satellites. Notice, however, that these observations served more as replies to objections against Copernicanism than as direct support of it (cf. the conclusions of *Siderus*). Yet, in December 1610, Galileo observed the phases of Venus, which he reported on December 30th in a letter to Cristoforo Clavio, concluding that the phases demonstrate that the planetary system is 'with no doubt different from the one commonly accepted', referring to Ptolemy's geocentrism. Of course, Galileo was well aware that that system is incompatible with the inner planets showing phases, but he took the phases of Venus not only as a refutation of Ptolemy's geocentrism but also as a 'strong argument for the Copernican arrangement', in his own words in a letter addressed to Kepler (quoted in Drake 1990, 137–138). According to Drake, it is only after this discovery that Galileo starts 'to speak out openly in favor of the Copernican system' (op. cit.
Q4 142) and this discovery 'sufficed to induce Galileo to make public his own Copernican
Q5 preference' (ibid. 145). Even Kuhn acknowledges that only these observations of the phases of Venus 'provide sufficient direct evidence for Copernicus' proposal' (Kuhn
Q6 1957, 222). Were Galileo, and his followers, justified in taking the phases of Venus not only as a refutation of Ptolemaic geocentrism but also as a 'strong argument' (precisely what we call Conf1-confirmation) for Copernicanism? To answer this question, we need a more detailed consideration of the context of this episode. At first sight, it could seem

that they were *not* so justified, since, as is well known (and as Kuhn and Drake emphasise when commenting on Galileo's taking of the phases of Venus as a strong argument for Copernicanism), heliocentrism was not the only system that predicted the phases of Venus. Indeed, Tycho Brahe's system, with the Earth at the centre, the Moon and the Sun orbiting around it, and the other planets orbiting around the Sun, also predicts them;[5] a fact that Galileo could hardly have failed to know (see Drake 1990; Dawes 2016). The key, to us, lies in whether Galileo did, or did not, take Tycho's system to be a viable rival to Copernicanism at that time.

Tycho died in 1601, and although his work on comets, parallax, and other topics, was well known and discussed, and his observations were still in use by many astronomers, such as his pupil Johannes Kepler, his planetary system was not seriously considered as an alternative worth considering; at least not by Galileo. In the *Dialogo sopra i due massimi sistemi del mondo* (1632), where he refers to Tycho many times in relation to comets, parallax and other issues, Galileo does not mention Tycho when Salviati comments on the phases of Venus. Moreover, in *Il Saggiatore* (1623), Galileo claims that Ptolemy and Copernicus provided complete systems of the world, something 'that I do not see Tycho has done' and a few lines later, he refers to Tycho's as a 'null' system (1623, §6). A few years before, in 1619, the influential Jesuit Horatio Grassi, had published his *Disputatio astronomica*, in which he defended Tycho's theory of comets: a defence rebutted the same year on Galileo's behalf by his friend and pupil Mario Guiducci, in a work entitled *Discorso delle comete*. In that pamphlet, Guiducci/Galileo delivers not only a reply about comets but also a devastating attack on Tycho's planetary 'system' for its absurdities and incongruities, such as its crossing planetary orbits. The pamphlet increased the belligerence of the Jesuits against Galileo (Dawes 2016) and was immediately attacked by Grassi in his *Libra astronomica* under the pseudonym 'Lottario Sarsi', the character to whom Galileo replies in *Il saggiatore*, where, as we have indicated, he disregards Tycho's system as a proper system of the world worthy of consideration.

This story reveals the reason why Galileo and his followers could take the phases of Venus as a 'strong argument' for Copernicanism despite the fact that they were also predicted by Tycho's system, a fact of which they were aware, as we have seen. It seems clear that they simply do not take Tycho's 'system' to be a serious rival hypothesis. Tycho's system was, in the aforementioned context, just a refuge for Aristotelian Jesuits, who resurrected it as an alternative in order to defend geocentrism from the difficulties that the phases of Venus posed for standard, Ptolemaic geocentrism (Solís and Sellés 2013). It is also clear, on the other hand, that the preference of Aristotelians for Tycho's system over Copernicus' was not due to Tycho's successful prediction of the phases of Venus, but to independent (mainly religious) reasons. Had Galileo considered Tycho's (or Wittich's) model to be a serious astronomical rival, he could hardly also have considered the phases of Venus to be a *strong reason* for accepting Copernicanism, as he actually did.

We think this piece of history supports the distinction between Conf1 and Conf2 and the idea that, in many contexts where acceptance, or belief, or action, are involved, it is Conf1 which is in play. For Galileo, the observation of the phases of Venus provided grounds (a 'strong argument') to accept heliocentrism as the true system of the world, and did not merely increase its plausibility; and the fact that he no longer deemed the Tychonic system a viable hypothesis seems crucial to his reasoning. Of course, one might simply deny that Conf1, and its associated argument (CONF), is ever at work or exists at all. We believe that historical episodes such as this, show this not to be the case; but we will not argue for this conclusion here in more detail. The ensuing discussion is, then, addressed and we hope of interest to those who (like Earman and Salmon, Giere, and many others) accept that something like (CONF) plays a role in relevant episodes of hypothesis testing. So, we will now proceed with our assessment of the precise content of RISK and its possible formalisation.

3. Riskiness, Unlikelihood, and Alternative Hypotheses

As we have already said, though (DER) suffices for refutation when P does not obtain, it is not enough to confirm a hypothesis (in the sense of Conf1) when P does obtain. To illustrate this, we have mentioned the situation in which H (together with A and IC) logically implies P, and P obtains, but there are other viable rival hypotheses considered in the context that also predict P. In this case, we would not say that H is confirmed, since the occurrence of P does not favour H against the rival hypotheses that also predict P.

As suggested above, examples like that of Galileo and the phases of Venus illustrate an important fact about the requirement that the confirmatory prediction not be predicted by other, incompatible viable rival hypotheses. As the qualification 'viable rival' connotes, this condition is *contextual*: the same prediction is not made by other rival hypotheses *taken into consideration in the given epistemic context*. So, the condition is highly pragmatic. It does not say that there are no other hypotheses that make the same prediction, which would never be the case.[6] What the condition demands is that the scientific context does not take *at that time* other hypotheses with the same prediction as worth considering, either because none has (yet) been formulated, or because any that have been formulated, once taken into consideration, are not deemed viable anymore, as in Tycho's case. Therefore, this condition is not just a matter of logic, like condition (DER), but contextual and temporal: it may be satisfied at one historical moment, but may not be satisfied later or earlier.

The condition that P is not predicted by alternative hypotheses in the given context is an interesting aspect of the risky condition. But there is another, more immediate aspect that is also essential and that we have already pointed out: that P is not very likely to obtain on its own, without the support of H. If P is a phenomenon that is highly expectable in and of itself, regardless of H or of some other rival hypothesis, then we would not say that H is confirmed by P being the case either. So, uniting these two aspects, in order for H to be confirmed by the occurrence of P, P should be a 'risky prediction', that is, an observable fact that is *both* (i) rather unexpected on its own *and* (ii) not known in the context to be predicted by alternative viable hypotheses. The informal formulation (RISK) above attempts to capture precisely this: that the prediction should be risky, and that the riskier the prediction, the stronger the confirmation it can confer.

It is not easy, however, to formalise this idea and make its logical form explicit. The condition is present, implicitly or explicitly, in many works and textbooks; we will mention just a few. Carnap's (1950, 1960) different measures of confirmation increase with the unlikelihood of the evidence, which amounts to assuming something similar to (RISK) (although not making it explicit that evidence must not be predicted by other hypotheses in the context). Similar considerations apply to Popper's measures of the severity of a test.

Popper (1963/1978) defines this measure as a function, $S(\mathrm{E,H})$, of the hypothesis (H) and the evidence (E) that the test is meant to ascertain, in two different ways (that covary): $S(\mathrm{E,H}) = \Pr(\mathrm{E|H}) - \Pr(\mathrm{E})/\Pr(\mathrm{E,H}) + \Pr(\mathrm{E})$ (whose denominator is posed for technical convenience and whose numerator embraces all the intuitive content, according to Popper himself); or $S'(\mathrm{E,H}) = \Pr(\mathrm{E|H})/\Pr(\mathrm{E})$. If the evidence is entailed by the hypothesis, then we have $\mathrm{P(E|H)} = 1$ and the test is more severe the more unlikely the evidence is. So, using our terminology, the riskier the prediction is, the better the test would be.[7] Giere (1979) formulates the condition literally as follows: 'If [Not H and IC and A] then very probably Not P' (92). This is basically our informal phrasing above, which does not make the precise logical form explicit. In later work, Giere, Bickle, and Maudlin (2006) mention as a condition for confirmation that 'the prediction is not likely to agree with the data if the model does not provide a good fit to the world' (35); which again is similar to our (RISK) above, now in terms of models. Mayo has also discussed similar conditions in several publications (1991, 1996, 2010, 2018). For instance, she states her *severity criterion* as follows: '(SC): *Severity Criterion*: There is a very high probability that test T would not yield such a passing result, if T is false' (1991, 529; see also 2010, 22), which again is similar to our informal version above (RISK). In all these cases, if the RISK-like condition obtains, and the prediction is a success, then the hypothesis is confirmed, or corroborated, or validated, or whatever term one prefers; which amounts to accepting something like CONF as the ampliative justification behind this.

These authors, though, make no explicit mention of the condition that no other hypotheses *at play in the context* make the same prediction.[8] An explicit mention of this so-called 'problem of alternative hypotheses', related to our condition, is made by Earman and Salmon:

> This fact poses a profound problem for the hypothetico-deductive method. *Whenever an observational result of an H-D test confirms a given hypothesis, it also confirms infinitely many other hypotheses that are incompatible with the given one.* In that case, how can we maintain that the test confirms our test hypothesis in preference to an infinite number of other possible hypotheses? This is the problem of alternative hypotheses. (1992, 49, their emphasis; notice that, in this quotation, the first two instances of 'confirm' refer to Conf2 while the third refers to our Conf1)

Clearly, here Earman and Salmon are referring to the formal fact that any dataset that fits a given hypothesis also fits infinitely many other incompatible hypotheses. But, as we have already mentioned, (RISK) is not violated merely by this formal fact; for then it would *always* be violated and no hypothesis would ever be confirmed (Conf1-confirmed)! In our Conf1 sense, (RISK) is violated if in the specific testing context there are other alternative hypotheses that make the same prediction *that are taken into consideration*. And, as our historical example in Section 2 illustrates, this is not always the case. Let us now inspect whether the logical form of this risky condition, including the absence in the given context of viable rivals making the same prediction, can be made more precise.

4. Riskiness as a Material Implication

The first, simpler option to explore seems to be to read (RISK) as a material conditional, that is, a material implication with antecedent '$\neg \mathrm{H} \wedge \mathrm{A} \wedge \mathrm{CI}$' and consequent 'very likely $\neg \mathrm{P}$':

$(\text{RISK})_m \quad \neg H \wedge A \wedge CI \rightarrow$ very likely $\neg P$

Recall that, as we have indicated above, it should be possible to decide whether both (DER) and (RISK) obtain before performing the test. Yet, contrary to what happens with the material reading of (DER), the truth of $(\text{RISK})_m$ *is* knowable before performing the test. This is so because $(\text{RISK})_m$ is true if its consequent is also true, and whether 'very likely ¬P' is true may be ascertained independently of the actual truth or falsehood of P. So, at least in this respect, $(\text{RISK})_m$ seems to be a candidate worth considering for explicating the logical form of (RISK).

There are nevertheless other reasons why it is difficult to read (RISK) as a material conditional, which is true whenever its consequent is true, regardless of what happens with the antecedent. In order to see why, suppose (as we think actually is the case) that the appearance of a new luminary in a specific region, X, of the sky in late 1758 was very unlikely and not predicted by any hypothesis in a given context. Then, the fact that no new luminary appears in X in late 1758 is very likely. It follows that the ensuing material conditional

(1) If Fresnel's Law is false, then, very likely no new luminary will appear in late 1758 in region X of the sky

is also true, because its consequent is indeed true. Assume now, as happened, that in fact a new luminary does appear in late 1758 in region X of sky, and consider the following argument:

(Fresnel) If Fresnel's law is false, then, very likely no new luminary will appear in late 1758 in a region X of the sky
A new luminary appeared in late 1758 in region X of the sky

Fresnel's law is true

If (Fresnel) were a valid inductive argument, given that its two premises are true, the conclusion would be established with high probability. But, of course, nobody would accept that the appearance of a new luminary in the sky confirms Fresnel's laws. So, clearly, (Fresnel) is *not* a valid inductive argument.

Now, if in (CONF) we omit A and IC (as we are going to do hereafter, since nothing in what follows hinges on this simplification) and we read (RISK) as $(\text{RISK})_m$, we obtain:

$(\text{CONF})_m$

not-H → very likely not-P $(\text{RISK})_m$
P

H

Notice that (Fresnel) is an instance of $(\text{CONF})_m$. Therefore, given that (Fresnel) is not a valid inductive argument, neither is $(\text{CONF})_m$. So, this shows that $(\text{RISK})_m$ cannot be an adequate analysis of the risky condition, since, by assumption, *the correct reading of (RISK) is precisely one that does make (CONF) a valid inductive argument*. We can therefore conclude that the risky condition is not properly formalised as a material conditional.[9]

In order to conclude that the risky condition is not a material implication, we have granted that (CONF) is the correct form of the confirmatory argument; but one might argue that this is far from clear. Riskiness cannot be a material implication if the confirmatory argument has only one other premise, namely, the occurrence of the prediction. In such a case, the problem seems to be that a positive connection between H and P is lacking. This may be taken as an indication that one can have a valid confirmatory argument with $(\text{RISK})_m$ as one premise but then adding something that expresses a positive connection between H and P. What could this connection be? The most natural candidate is the derivability condition (DER). As we mentioned above, this seems to be the option taken in Giere (1979, 99), where it is claimed that, in order to inductively conclude H from P and 'if not H then very probably not P' (recall that for the sake of simplicity we are now omitting A and IC), 'one must first have a good test. And that requires Condition 1 as well as Condition 2'; where his Condition 1 is our (DER) and his Condition 2 our $(\text{RISK})_m$.[10] So, following this route, the inductive confirmatory argument would have the following form:

(CONF-DER_m)

not-H → very likely not-P (RISK_m)
H entails P (DER)
P

H

However, *pace* Giere, this is *not* a valid inductive argument either. The first and the third premises may be true just given that a very unlikely P obtains. The second is true when P is predicted by H. So, the three premises are true if H predicts an unlikely P that is nevertheless true. But *this* does *not* need to render H highly probable because, in the same context, there can be other incompatible hypotheses that make the same prediction. So, once again we have come across the problem of alternative rival hypotheses. The following example illustrates this point. As background knowledge, let us assume that, on 30th November 2020, John is a poor novelist with almost no savings yet a beloved nephew of an old, rich and very sick aunt who is, in addition, a childless widow. However, rather surprisingly, John ends the year 2020 with €1M in his bank account. This last fact counts as our prediction (P). Now, there is a hypothesis, H, that entails this prediction: John's bank has decided to give €1M to every customer named John on 31st December 2020. Here, the material implication 'If John's bank does not decide to give €1M to every customer named John on 31/12/2020, then very likely John will not end 2020 with €1M in his account' is true, basically because the consequent is true in this context. In addition, H implies P. So the three premises of the argument are true, yet we should not accept it as highly probably that John's bank has in fact decided to give €1M to every customer called John on 31/12/2020. There is, in this context, another hypothesis that would account for the unlikely fact that John ends 2020 with €1M in his bank account: his aunt suddenly dies, leaving him the money.

Notice that we are not arguing simply that our hypothesis may be false, since this is of course always the case in *good* inductive arguments. What we defend is that adding (DER) to (RISK_m) does not generate a good inductive argument. The reason, as we have seen, is that one is *not* inductively (fallibly) justified in believing H in cases in which, even if (DER) is satisfied, there are other viable hypotheses that deserve consideration in the context that also imply P. So, to conclude, not even adding derivability as an additional premise does the confirmation argument work with riskiness reconstructed as a material implication. And we do not see any other way to make the material reading of riskiness work. Now, if the logical form of (RISK) is not a material conditional, what is it? It seems that a plausible way to proceed is to analyse it in terms of conditional probabilities. We will next explore the prospects of such analysis.

5. Riskiness and Conditional Probabilities

The most straightforward candidate for analysis of (RISK) in terms of conditional probabilities is ‘Pr(¬P|¬H) is high’, which simply translates our initial informal appraisal of the risky condition into the language of conditional probabilities. With this analysis, the confirmatory argument (CONF) reads as follows (for simplicity, we continue to omit A and IC):

($RISK_{prob}$) Pr(¬P|¬H) is high
P

H

This argument, however, is not valid. Pr(¬P|¬H)≈1 does not imply Pr(H|P)≈1 (take e.g. P as being left handed and H as being male) and, therefore, from both Pr(¬P/¬H)≈1 and P, it cannot be deductively derived that Pr(H)≈1, so neither can H be inductively derived with a high probability. This shows that, if (CONF) is a valid ampliative argument, its first premise (RISK) *cannot* be formalised as ‘Pr(¬P|¬H) is high’. Is there any better way to analyse (RISK) in terms of conditional probabilities? One option might be to maintain ‘Pr(¬P|¬H) is high’ as part of the analysis and think of what we could add to it in order to guarantee ‘Pr(H|P) is high’. The following considerations seem to suggest one candidate.

Consider the following argument (for simplicity’s sake, we continue to omit A and IC when they are unnecessary):

(SYP)
If the mayor does not have untreated syphilis then very likely he does not have paresis
The mayor has paresis

The mayor has syphilis

Here, we are assuming that untreated syphilis is practically the only condition that can lead to the development of paresis but that not *all* individuals that have untreated syphilis Q7 develop paresis; in fact, roughly only a quarter of them do.[11] So, in terms of conditional probabilities, we have that Pr(paresis|¬ untreated syphilis)≈0 and Pr(paresis|untreated syphilis)≈0.25 and, therefore, Pr(paresis|untreated syphilis) >> Pr(paresis|¬ untreated syphilis).

The first thing to notice is that (SYP) is an instance of the confirmatory argument (CONF) (with H ≡ ‘The mayor has untreated syphilis’ and P≡ ‘The mayor has paresis’) *if*, of course, the ‘if … then-very-likely … ’ in the first premise of both arguments is interpreted in the same way. Now, given the background information that we have just provided and considering that no other source of paresis apart from untreated syphilis is known in the context, nobody would doubt that (SYP) is a valid inductive argument and that the mayor's paresis confirms that he has untreated syphilis. So perhaps we can learn what makes (CONF) a valid inductive argument by considering why (SYP) is one.

The comparison between the arguments allows us to arrive at some morals. First, to the extent that it is accepted that the mayor's paresis (P) confirms that he has untreated syphilis (H), it should be clear that (DER) is not required in order for the confirmation because in the (SYP) case, clearly, one cannot derive the mayor's paresis from the fact that he has untreated syphilis. This answers in the negative the question about the 'intriguing' relation between (DER) and (RISK) that we posed in the first section. Now, strict prediction, that is, logical derivability is the strongest form of positive connection between H and P, but there may be other, weaker forms. This is indeed what happens in the (SYP) case, where paresis is not predicted by untreated syphilis in the sense of being logically implied by it; nevertheless, the two are positively connected since the probability of having paresis given untreated syphilis, even if not high, is much greater than that of having it without untreated syphilis, which is close to 0. Thus, we still have a positive but not 'maximal' connection between H and P which, in terms of conditional probabilities can be expressed as $Pr(P|H) >> Pr(P|\neg H)$.

This consideration, in turn, can motivate an analysis of riskiness that adds the condition '$Pr(P|H) >> Pr(P|\neg H)$' to the clause '$Pr(\neg P|\neg H)$ is high', which we have formerly discussed. So, according to this proposal, riskiness is analysed in terms of conditional probabilities as follows:[12]

(RISK*) (i) $Pr(\neg P|\neg H)$ is high and (ii) $Pr(P|H) >> Pr(P|\neg H)$

The second subcondition is positive probabilistic relevance;[13] and recall that the two together are meant to capture the idea mentioned earlier that a hypothesis, H, cannot be confirmed by positive evidence, P, if there are other hypotheses in the context that also predict (or, more generally, that are also positively correlated with) P. Both conditions are satisfied in good confirmatory episodes, such as syphilis/paresis or the reappearance of Halley's comet; but either would fail in non-satisfactory confirmatory episodes, for instance (i) fails in the case of the phases of Venus (in the 1590 context, so when Tycho's hypothesis with the same prediction was taken into account) and (ii) fails in the (Fresnel) case.

Unfortunately, this new proposal to formalise riskiness does not work. As the following diagram shows, (RISK*) does not guarantee that there are no other hypotheses predicting P, nor that $Pr(H|P)$ is high (Figure 1). **Q8**

One reaction to this failure could be to think that, no matter what our initial intuitions might suggest, condition (i) stating that the conditional probability $Pr(\neg P|\neg H)$ is high does *not* serve to capture riskiness: not even if some other conditions are added to the analysis. Then, of course, we should attempt to replace (i) by some other condition. One option could be '$Pr(\neg H|P)$ is low', the idea being that the lower $Pr(\neg H|P)$ is, the smaller the space for other hypotheses *given* P. If each $Pr(H_i|P)$ is taken to be significant ($>>0$), then the existence of other H_i that make the same prediction would preclude satisfaction of '$Pr(\neg H|P)$ is low'. But again, this does not seem to work. For one thing: the disjunction of all such hypotheses satisfies this criterion, so the disjunction would be

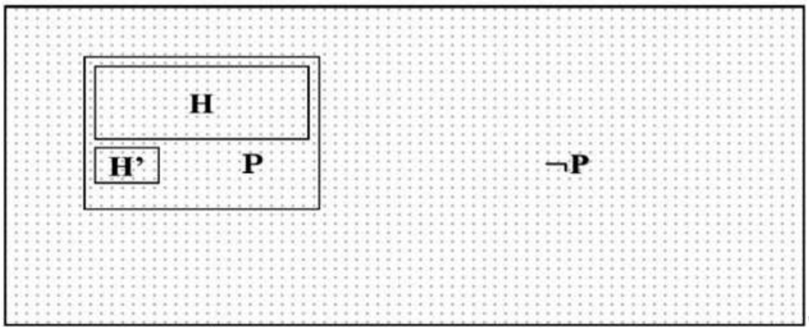

Figure 1. Here, both H and H′ entail P. This is so because all points that are in H are also in P, and the same occurs with all points that are in H′. Additionally, Pr(P|H)>>Pr(P|¬H), since Pr(P|¬H) is low and Pr (P|H) = 1. Pr(¬P|¬H) is also high, since the overwhelming majority of points that lie in the complement of H are not in P. Therefore, H satisfies (RISK*) yet this does not exclude there being another incompatible hypothesis, H′, that also entails P. Notice, in addition, that all the same things can be said of H′, so nothing hinges here on whether the initially considered hypothesis or the putative rival(s) has a higher absolute (prior) probability.

confirmed. Yet, although in some very specific contexts a disjunctive hypothesis may be worth considering, in general it is not the case that theoretical scientists, technologists, physicians, attorneys, detectives and the like consider disjunctive hypotheses to be acceptable; so this is not a general way of Conf1-choosing a hypothesis. Another more important consideration is that 'Pr(¬H|P) is low' is precisely equivalent to 'Pr(H|P) is high'; but, as we have been suggesting, this is intuitively similar to the very idea of P inductively supporting H, as intended in (CONF). This depends on how we read this probability; but if it is read contextually as riskiness, then this seems to be the *explicandum*, not the *explicans*. If, alternatively, it is not read in this way, then it can hardly exclude other hypotheses that make the same prediction.

6. A Threshold Condition for Riskiness

Another kind of reaction could be to claim that the reason why the subcondition 'Pr(¬P|¬H) is high' does not work as a formalisation of riskiness (not even when a positive correlation between H and P is added) is that it is too unspecific, and so some more specific alternative could work. For instance, that Pr(¬P|¬H) must be much higher than Pr(¬P), that is, that the falsehood of H makes P considerably less probable, in a sense of 'considerably less probable' to be specified. One such specification could be to demand that ¬H decrease the initial probability of P by more than a half:

(RISK**) Pr(P|¬H) / Pr(P) < 1/2

This proposal seems to be well motivated if one consider (as Cruppi and Tentori 2010, 2012, for independent reasons, do) that a good measure of confirmation is [Pr(H|P)−Pr(H)] / Pr(¬H), which is equivalent to 1−[Pr(P|¬H) / Pr(P)]. In such a case, the smaller Pr(P|¬H) / Pr(P) is, the greater the confirmation and, as a consequence, the more ¬H decreases the prior probability of P, the more P confirms H. In this regard, a candidate for a natural confirmatory threshold could be decreasing the probability of

P by a half, which thereby could be interpreted as a formalisation of riskiness, since there Q9 cannot be two incompatible hypotheses that predict P and both satisfy this condition.[14] To see that this is the case, suppose that H_1 and H_2 are incompatible and that both satisfy (RISK**); then, by Bayes' theorem, $\Pr(\neg H_i|P)$ / $\Pr(\neg H_i) < 1/2$, and since $\Pr(\neg H_i) < 1$, it follows that $\Pr(\neg H_i|P) < 1/2$ and thus $\Pr(H_i|P) > 1/2$. Now, given that the two hypotheses are incompatible, we finally obtain that $\Pr(H_1 \text{Ú} H_2|P) = \Pr(H_1|P) + \Pr(H_2|P) > 1$. Contradiction.

Is (RISK**) is a good formalisation of riskiness? Unfortunately, we do not think so. Recall that riskiness conveys that P is unlikely on its own and not predicted by other (incompatible) hypothesis considered in the context. Thus, according to this informal reading of riskiness, for a probabilistic condition μ(H,P) to express riskiness, it should imply the following:[15]

(#) If H predicts P and satisfies μ, then there is no incompatible hypothesis, H', in the context that also predicts P.

It is clear, though, that (RISK**) does not imply (#): if H predicts P and ¬H reduces the probability of P by more than half, there can perfectly well be other hypotheses in the context that also predict P, as Figure 2 illustrates.

What probabilistic condition μ(H,P) on H and P alone, could imply, together with H predicting P, that there is no other H' that is incompatible with H (i.e. with $\Pr(H \wedge H') = 0$) that also predicts P? It is hard to think of an acceptable one. We say 'acceptable', since there certainly are such conditions, namely, any contradicting $\Pr(P|H) = 1$ (or its substitute for just a positive relation instead of strict prediction) and thus implying everything; such a condition would formally do, but of course it would not explicate anything. And we do not see any other.

One may argue that there is no such viable condition just because (#) is a too strong a demand, given that its consequent does not impose that H' also satisfy μ. If we added this to the consequent, we would have that μ captures riskiness if it implies the following:

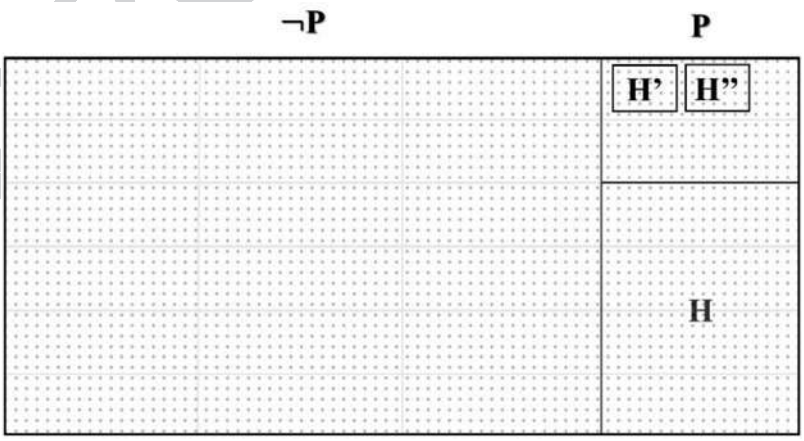

Figure 2. Here $\Pr(P|\neg H) = 1/10$ and $\Pr(P) = 1/4$. So $\Pr(P|\neg H) / \Pr(P) = 2/5 = 0.4$ and $\Pr(P|\neg H) / \Pr(P) < ½$ holds. However, H' and H'' also imply P (and there could be infinitely more).

> (##) If H predicts P and satisfies μ, then there is no incompatible H' in the context that also predicts P *and* satisfies μ.

Now, there are such conditions (different from merely contradicting 'H predicts P'), and (RISK**) is precisely (the most intuitive?) one: since there cannot be two incompatible hypotheses H and H' that both satisfy (RISK**), then *a fortiori* (RISK**) implies (##). Does this mean that condition (RISK**) is a proper formalisation of riskiness? No. Firstly, we maintain that for a probabilistic condition μ(H,P) to be a candidate for expressing riskiness, it must imply (#), not just (##). But, secondly, even those who do not agree with us and defend the idea that implying the weaker (##) suffices, cannot take just *any way* of implying (##) as an adequate expression of riskiness. The reason why (RISK**) implies (##) is just that it cannot be satisfied by any two incompatible hypotheses, regardless of their capacity to predict P;[16] so, the reason why it implies (##) is completely independent of whether those hypotheses predict P or not, which is (part of) what riskiness is about. Generalising: those who accept that implying (##) suffices for a given μ to capture riskiness, must restrict themselves to μs that imply (##) not merely because it is impossible for μ to be satisfied by two incompatible hypotheses. So, whoever wants to defend that implying (##) is a criterion for riskiness, must propose a μ that can be satisfied by more than one incompatible hypothesis, which is not the case of (RISK**). We then could not come up with any probabilistic condition μ(H,P) that implies (#), nor any that implies (##) while being able to be satisfied by two or more incompatible hypotheses.

Defenders of (RISK**) may reply that it is unfair to ask them to imply (##) with a μ that can be satisfied by more than one hypothesis; for we are looking for a formalisation of a condition, riskiness, that tells us that no other hypotheses in the context make the same prediction. And this is a condition that is only satisfied by one hypothesis, such as the condition (RISK**) they propose. This is fair enough, but cannot imply that any condition that can be satisfied (in a context) by just one hypothesis expresses that no other hypothesis in the context makes the same prediction; and (RISK**) actually does not, as Figure 2 illustrates. Notice that riskiness is satisfied, for instance, in 1610 for heliocentrism and the phases of Venus but not, say, in 1590; although, upon a standard understanding of probability, (RISK**) is satisfied or not in both contexts alike. Now, our detractors may again reply that this is not necessarily so, that the meaning of 'probability' in (RISK**) is such that it makes (RISK**) true in 1610 but false in 1590. This is fair, again; but then the sense of 'probability' they are now using is irreducibly pragmatic and we do not see what explicatory role can it play: How would they assess whether $Pr(P|\neg H)/Pr(P)<½$ is false in the 1590 context? (Recall that one must know whether riskiness obtains in order to assess a testing episode.) We do not see any other way besides assessing directly whether the prediction was made by other hypotheses considered in the context, in which case the probabilistic apparatus plays no elucidatory role at all and is just a pseudo formal rephrasing of riskiness.

7. Riskiness and Comparative Probabilities

Before despairing of a probabilistic characterisation, there is one last option for reconstructing riskiness in terms of conditional probabilities that we wish to consider, yet it does not refer exclusively to H and P but rather makes explicit reference to other possible

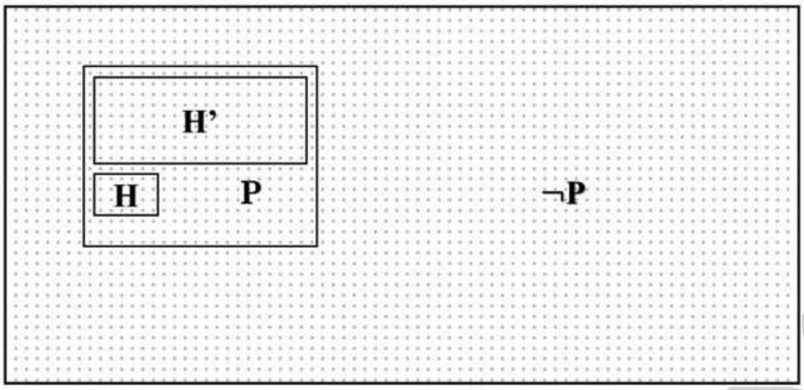

Figure 3. Here, both H and H′ entail P, and nevertheless: Pr(P|¬H′) ≤ Pr(P|¬H).

hypotheses in the context. A trivial option would be ‘there is no other H’ considered in the context such that Pr(P|H’) = 1’. This obviously works since Pr(P|H’) = 1 means that H’ also predicts P, and the condition that there is no other such H’ considered in the context is simply the very claim that there are no known rivals in the context that imply the same prediction. All we have done here is some reworded, writing ‘Pr(x|y) Q10 = 1’ instead of ‘y implies x’, which of course is not explicatory at all (the same applies if instead of strict prediction we talk of a relevant positive relation, i.e. ‘Pr(P|H’) is high’, or even merely of ‘Pr(P|H’) >> Pr(P|¬H’)’. Is there any other analysis, with some explicatory import, referring to alternative hypotheses entertained in the context? One candidate could be the following condition:

(RISK***) for every other H’ considered in the context: Pr(P|¬H’) > Pr(P|¬H)

The idea is that, if H does but H’ does not predict P, then H is more closely linked to P than H’ is, and denying H negatively affects P more than denying H’ does; therefore, P is more unlikely given that not H than given that not H’. If this is correct, then when there are no rivals in the context that make the same prediction, the condition is satisfied; and when there are such rivals, it is violated. However, contrary to the motivating idea just mentioned, the fact that H entails P does not preclude the possibility of another hypothesis, H’, taken into consideration in the same context, that also predicts P and is such that Pr(P|¬H’) ≤ Pr(P|¬H), as Figure 3 illustrates.

So, in the absence of a good candidate for this last strategy, we provisionally conclude again that there is no good (i.e. elucidatory) probabilistic formalisation of riskiness in terms of H, P and referring to other hypotheses in the context.

8. Concluding Remarks

To summarise the discussion of the logical form of the riskiness condition, neither a material conditional with a probabilistic consequent, nor an analysis using conditional probabilities seems adequate to capture riskiness. We have not been able to come up with a different, satisfactory formal analysis of riskiness, and given the highly contextual and pragmatic character of this condition, we are pessimistic that this can be done. The likelihood/probability in (RISK) is highly pragmatic. Of course, one may *define* a notion of ‘confirmatory pragmatic probability’ (‘Pr*’) for a prediction meaning ‘P is not very likely in itself, nor predicted by other acceptable rival hypotheses in the context’, and then use Pr* to ‘formalize’ (RISK). But this would be an empty explication (notice that if the probability in (RISK) has this pragmatic unanalysable character, then the ‘high probability’ of the inductive force of (CONF) inherits this peculiar pragmatic character). In the absence of new proposals with better prospects, we conclude that (RISK) has a primitive, ineliminable pragmatic nature that resists any formalisation using more basic logic or probabilistic terms.

With regard to whether derivability is needed for confirmation, derivability links prediction and hypothesis in a *positive* way, constraining the status of the prediction when the hypothesis is true. The risky condition, as we have phrased it informally, links hypothesis and prediction in a *negative* way, since it constrains the status of the prediction if the hypothesis is not true. Arguments like (Fresnel) show that there must be a non-contingent relation between hypothesis and evidence in order for the evidence to confirm the hypothesis. Arguments like (SYP) show that this non-contingent relation must be ‘positive’; nevertheless, as (SYP) also shows, this positive connection may be partial: it is possible to confirm a hypothesis when the evidence obtains, even if (DER) does not hold, i.e. even if the evidence is not strictly implied by the hypothesis, provided, as we have said, that there is at least a non-contingent positive relation between H and P. So, is (DER) dispensable for confirmation?

First of all, of course (DER) is not dispensable for testing in general, if testing is designed to confirm *or* disconfirm, since for *deductive* disconfirmation/refutation (DER) is necessary, and no weaker condition would do. What about confirming scenarios? As we have just said, arguments like (SYP) show that it is possible to confirm a hypothesis when P obtains even if (DER) does not hold, provided that there is a non-contingent positive relation between H and P. So, (CONF) works only if there is an implicit premise telling us that there is a positive non-contingent relation between H and P: in the case of strict prediction, this non-contingency is deductive implication, whereas in weaker cases like (SYP), the non-contingency amounts to a significant increase of probability (not reaching 1, though). So, is this non-contingent positive relation between H and P already part of the content of (RISK)? We submit that this is indeed the case and that riskiness includes these three clauses:

R0: H is non-contingently positively correlated with P.

R1: P should be rather unexpected/unlikely on its own.

R2: P should not be known to be predicted by alternative rival hypotheses considered viable in the context.

This would suffice in general for (CONF).

Let us conclude with a caveat on the significance of our results. One may argue that if, to start with, confirmation understood as Conf1 has an ineliminable pragmatic or contextual character, then it does not come as a surprise that it cannot be formalised. Yet not everybody that accepts that (CONF) has a role in the confirmation of scientific hypotheses also acknowledges that confirmation has such a contextual character, see for instance the reference to Giere above. Therefore, noticing that the acceptance of

(CONF) commits us to accepting that confirmation is contextual is a point worth making.

If we are right in claiming that confirmation—in the sense of Conf1—is irreducibly contextual, then the notion of confirmation is similar in this regard to others, such as that of explanation or representation, with components that are essentially pragmatic. This does not constitute a devaluation of the notion or discredit it, but simply shows the limitation of philosophical attempts to understand important concepts in science using purely formal methods. In addition, the fact the confirmation is contextual does not mean that external, non-epistemic elements (such as aesthetical or ideological values) necessarily play a crucial role in the confirmation of scientific hypotheses. Here, we have not dealt with the question of what determines the contextual assessment of the viability of alternative hypotheses, yet that assessment can perfectly well be grounded in epistemic reasons, as our discussion of the case of the phases of Venus may suggest. This important issue, though, goes beyond the scope of the present paper.

Notes

1. Here, we do not need to assume that the prediction P is strictly speaking a purely observational claim, nor that there is a sharp observational/theoretical divide. The only requirement is that there be a relevant contrast between the hypothesis and the prediction with regard to empirical assessment such that an experimental determination of whether the latter occurs is substantially more direct.
2. For an explanation of what the auxiliaries or the initial conditions are, see, for instance, Giere (1979, 87–88).
3. Actually, one may not know it after the test either, since it is possible that P fails, in which case knowing that the material conditional is true would require us to know that something in $H \wedge A \wedge IC$ failed, which we may not know—of course we know this if we know that P failed *and* we also know that the conditional is true.
4. We thank … for this observation.
5. Tycho's was not the only geocentric system predicting the phases of Venus. Although not as well known, Paul Wittich's system (inspired by Martianus Capella's work, and most probably inspiring Tycho's system), in which Mercury and Venus orbit around the Sun, and the Sun, the Moon and the other planets orbit around the Earth at the centre, also predicts the phases of Venus.
6. Notice, in this respect, the trivial logical fact that, for every hypothesis, H, and prediction, P, made by H, there conceptually exist infinitely many other incompatible, alternative hypotheses that make the same prediction. In addition, in most cases it is possible to concoct alternative—often quite artificial—hypotheses with the same prediction.
7. With this caveat, we only aim to show that Popper's measure of the severity of a test captures one aspect of riskiness, not that it is equivalent to this latter condition. Moreover, given Popper's explicit anti-inductivism, we are perfectly aware that he would *not* endorse an argument like (CONF). However, it remains an interesting question whether Popper's 'severity of a test'—or his closely related 'degree of corroboration'—should be considered *pace* the author an inductive measure (see Díez 2010 for a discussion of this issue).
8. Mayo (1996, ch. 6) addresses the issue of alternative hypotheses, but in the context of the general problem of underdetermination of theory by data, explicitly referring to all conceptually possible alternatives, not only to the contextually considered rivals, which differs from the problem we deal with here.
9. In general, the argument '(i) not-H → very likely not-P, (ii) P, THEREFORE (iii) H' is neither deductively nor inductively valid. It is not deductively valid since the second premise is not the negation of the consequent of the first premise (the occurrence of P is

compatible with its unlikelihood). And it is not inductively valid either, since the two premises may be true due to the actual occurrence of a very unlikely P, which tells us *nothing* about the possibly completely unrelated H (as in the (Fresnel) case above).

10. Giere claims that the conclusion in point is not literally ‘H’, but ‘approximately H’. This is not intended to substitute the inductive argument for a deductive one, but to qualify the conclusion of the inductive argument. The reasons Giere provides for such a move are worth considering; but since nothing of what follows hinges on them, we will not discuss them here.
11. Cf Salmon (1980, 49).
12. We should add that Pr(P) is low, in order to capture the idea that P is unlikely on its own, but since nothing in what follows hinges on this, we omit it for the sake of simplicity.
13. Notice that we are not seeking a probabilistic explication of causation, in which case Pr (P|H) >>Pr (P|¬H) would not be a candidate for well-known reasons related with the existence of co-effects; we are just looking for non-contingent evidential relations that may perfectly well include common cause co-effects of the barometer/storm kind.
14. XXX, personal communication.
15. For the sake of simplicity, in what follows we talk of H predicting P; but as explained before, this could be relaxed to just a positive relation in the sense of condition (RISK*ii), which is compatible with, but may be weaker than, strict prediction.
16. Notice that to conclude that two hypotheses H_1 and H_2 cannot both satisfy (2*ii#), we have only assumed that they are incompatible and some axioms/theorems of probability theory, such as Bayes’ theorem.

Funding

This work was supported by Ministerio de Ciencia Tecnología y Telecomunicaciones [grant
Q2 number PID2020-115114GB-I00].

References

Carnap, R. 1950. *Logical Foundations of Probability*. Chicago: U. Chicago P.

Carnap, R. 1960. “The Aim of Inductive Logic.” In *Proceedings of the 1960 International Congress for Logic, Methodology and Philosophy of Science*, edited by E. Nagel, P. Suppes, and A. Tarski, Vol. 1, 303–318. Stanford, CA: Stanford U. P.

Conant, J. 1950. *The Overthrow of the Phlogiston Theory; the Chemical Revolution of 1775–1789*.
Q11 Cambridge, MA: Harvard Univ. Press.

Cruppi, V., and K. Tentori. 2010. “Irrelevant Conjunction: Statement and Solution of a New Paradox.” *Philosophy of Science* 77: 1–13.

Cruppi, V., and K. Tentori. 2012. “A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems).” *Philosophy of Science* 79: 365–385.

Q12 Dawes, G. W. 2016. *Galileo and the Conflict Between Religion and Science*. Taylor & Francis. p. 103.

Díez, J. 2010. “On Popper’s Strong Inductivism.” *Studies in History and Philosophy of Science Part A* 42: 105–116.

Drake, S. 1990. *Galileo: Pioneer Scientist*. Toronto: University of Toronto Press.

Earman, J., and W. Salmon. 1992. “The Confirmation of Scientific Theories.” In *Philosophy of*
Q13 *Science*, edited by M. Salmon et al., 42–103. Englewood: Prentice Hall.

Q14 Galileo, Galilei. 1964. *Opere*. Edited by F. Brunetti. Turin: Unione Tipografico-Editrice Turinese.

Giere, R. 1979. *Understanding Scientific Reasoning*. New York: Holt, Reinhart and Winston.

Giere, R., J. Bickle, and R. Maudlin. 2006. *Understanding Scientific Reasoning*. London: Thomson Nelson.

Q15 Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago, IL: Univ. Chicago Press.

Mayo, D. 1991. “Novel Evidence and Severe Tests.” *Philosophy of Science* 58: 523–552.

Mayo, D. 1996. *Error and the Growth of Experimental Knowledge*. Science and Its Conceptual Q16 Foundations Series. University of Chicago Press.

Mayo, D. 2010. "Learning from Error, Severe Testing, and the Growth of Theoretical Knowledge." In *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the* Q17 *Objectivity and Rationality of Science*, edited by D. G. Mayo and A. Spanos. Cambridge Q18 University Press. pp. 28.

Mayo, D. 2018. *Statistical Inference as Severe Testing*. New York: Cambridge University Press.

Popper, K. R. 1963/1978. *Conjectures and Refutations*. London: Routledge and Kegan.

Solís, C., and M. Sellés. 2013. *Historia de la ciencia*. Barcelona: Espasa.

Sprenger, J. 2011. "Hypothetico-Deductive Confirmation." *Philosophy Compass* 6/7: 497–508.

Q19 White, J. H. 1973. *The History of the Phlogiston Theory*. London: AMS Press.