

Predictivism and avoidance of ad hoc-ness: an empirical study

Samuel Schindler
Aarhus University

Abstract

Predictivism is the thesis that evidence successfully predicted by a scientific theory counts more (or ought to count more) in the confirmation of that theory than already known evidence would. One rationale that has been proposed for predictivism is that predictive success guards against ad hoc hypotheses. Despite the intuitive attraction of predictivism, there is historical evidence that speaks against it. As valuable as the historical evidence is, however, it is largely *indirect* evidence for the epistemic attitudes of *individual* – albeit prominent – scientists. This paper presents the results of an empirical study of scientists' attitudes toward predictivism and ad hoc-ness (n=492), which will put the debate on a more robust empirical footing. The paper also draws attention to a tension between the ad hoc-ness avoidance rationale of predictivism and the ways philosophers have spelled out the notion of ad hoc-ness.

Keywords: predictivism; prediction; accommodation; ad hoc hypothesis; confirmation; empirical study

1 Introduction

The successful prediction of new phenomena is widely considered to be important by philosophers of science. Lakatos (1978) for example once wrote that “the only relevant evidence [in the assessment of a theory] is the evidence anticipated by a theory” (p. 38). Popper made it a requirement for a theory that it “lead[s] to the prediction of phenomena which have not so far been observed” (Popper 1963/1978, 241). The view that a theory's successful predictions generally inspire *more* confidence in the theory than the theory's accommodations of already known phenomena is also known as the thesis of “predictivism”, or more precisely, “temporal predictivism”.¹ In one guise or another, predictivism has many contemporary supporters (Maher

¹ An extreme version of predictivism is that a theory gains *no* confirmation (or corroboration) from evidence that it ‘merely’ accommodates. Although the remarks by Lakatos and Popper suggest that they may have

1988, Lipton 1991/2004, Hitchcock and Sober 2004, Worrall 2014), is widely presumed to be true by many philosophers in the realism debate (Psillos 1999, Vickers 2013, Saatsi 2019)², and seems to be entailed by face-value readings of Bayes' theorem and resulting in the notorious "problem of old evidence" (Glymour 1980, Sprenger and Hartmann 2019). But is predictivism actually true? And if so, what is its rationale?

Just like many other questions in the philosophy of science, the debate about predictivism has a descriptive *and* a normative aspect: does predictive success actually have more weight in theory appraisal by scientists? And ought predictive success to count more than accommodative success? Both aspects are important: if the majority of scientists were to behave as if predictivism was false, then philosophical accounts presuming its correctness would fail to be philosophical accounts of actual science. There may be some ideal of how science ought to be that is never ever realized. And it may be interesting to speculate about that ideal. For example, philosophers have conducted *thought experiments* to determine a rationale for predictivism. A particularly widely discussed thought experiment involves the prediction of the next outcome in a series of coin tosses is also meant to not only lend credence to predictivism but which also is supposed to enlighten us about the right rationale of predictivism (Maher 1988, Lange 2001, Harker 2006).³ The evidential relevance of such thought experiments, however, remains unclear; coin-flipping is arguably quite different from what's at stake in the confirmation of scientific theories (Harker 2006).⁴

Other, more descriptively oriented contributions to the predictivism debate have conducted detailed analyses of scientists' epistemic attitudes toward a theory's successful novel predictions in selected historical episodes. Prominent historical case studies include the Fresnel white spot prediction (Worrall 1989), Einstein's successful prediction of light bending (Brush 1989), and Mendeleev's prediction of new chemical elements (Maher 1988, Lipton 1991/2004, Scerri and Worrall 2001, Schindler 2014). In Worrall's seminal discussion of the white spot prediction by the wave theory devised by Augustin-Jean Fresnel in 1818, for example, Worrall has argued against common historical narratives, and against our prior intuitions, that scientists actually did not regard Fresnel's successful prediction as superior evidence for the theory at all (Worrall 1989). Instead, they paid much more attention to Fresnel's successful accommodations of straightedge

had this extreme version in mind, this version has played no role in subsequent discussions. Note also that predictivism is not simply the thesis that theories receive more confirmation when they help extend the basis of known facts which they accommodate.

² Realists tend to commit to the approximate truth of only those theories, which managed to generate novel predictions.

³ The original thought experiment consists of tossing of a coin 100 times. Who would you trust more in predicting the outcome of the 100th toss: a person who observed the outcome of the first 99 times or a person who correctly predicted every single outcome of the first 99 tosses? Maher claims that most people would trust the second person's prediction of the 100th outcome more than the first person.

⁴ For one thing, what side of the coin will come up next is inherently unpredictable. In contrast, phenomena predicted by scientific theories are not inherently unpredictable.

diffraction patterns, which were long known, and which the emission theory of light had long struggled to explain.

The current contribution falls more on the descriptive side of the spectrum in the debate. But instead of analyzing yet another historical case study that provides indirect evidence for the epistemic attitudes of a few individual scientists, I will here present a systematic empirical study that seeks to *directly* get at the epistemic attitudes of the scientific *community* toward novel success. This has obvious advantages: not only does it avoid indirect inferences from the historical facts to the epistemic attitudes of scientists, but it also allows for statistically sound inferences and thereby avoids persistent questions about the representativeness that the case study approach has been plagued with (Chang 2011, Kinzel 2015).

The current paper also has a (fairly modest) normative goals. There is a link between the apparent attractiveness of predictivism and the avoidance of ad hoc accommodations. This link was perhaps first noticed by Worrall (1985; 1989), but contrary to what he believes, the “ad hoc-ness avoidance” rationale, as I will call it, actually favors temporally novel success. The paper also draws attention an uncomfortable tension between the ad hoc-ness avoidance rationale of predictivism and the ways in which philosophers have understood the notion of ad hoc-ness.

The study conducted for this paper surveyed both natural and social scientists (n=492), and philosophers (n=86) on their views of temporally novel success and ad hoc-ness. The main results of this study both contradict and expand the historical studies. Specifically, most natural scientists and philosophers expressed a general preference for temporally novel success, whereas social scientists do not. The majority of scientists and philosophers view ad hoc hypotheses as suboptimal, but only a minority categorically rejects their use. All three groups conceived of ad hoc-ness in terms of independent testability and parameter freedom.

The structure of the paper is as follows. In Section 2 I argue that there is a compelling argument for temporal predictivism, which has to do with the avoidance of ad hoc hypotheses. This view can already be found in Worrall’s work, although Worrall is in fact an opponent of temporal predictivism. The rationale is also fairly widespread in the literature, although perhaps not as overtly as it could be. Worrall believes that the ad hoc-ness avoidance rationale favors his own “heuristic account” and subscribes to what I call the epistemic symmetry thesis between temporally predicted evidence and certain kinds of accommodated evidence. Section 3 briefly reviews a different sense of ad hoc-ness that can be found in the literature and points out that all of these accounts imply the epistemic symmetry thesis, effectively undermining the ad hoc-ness avoidance rationale of predictivism. Section 4 describes the empirical study that was carried out for this paper. Section 5 discusses the results in depth and closes with a number of methodological remarks.

2 Predictivism and ad hoc-ness avoidance

Various forms of predictivism have been widely discussed in the literature. The two most important versions of predictivism are temporal predictivism and heuristic or use-novel predictivism; in what follows, I shall focus on these, because temporal predictivism is the “original”, most intuitive, and most straightforward predictivist thesis, and because heuristic predictivism has not only been the most influential and criticized predictivist thesis, but it also explicitly makes a conceptual connection between predictions and ad hoc-ness, which I think is worth exploring in more depth.⁵

Temporal predictivism is the most intuitive form of predictivism: predicted phenomena ‘count more’ in the confirmation of a theory than accommodations when the predicted phenomena are not known to the scientific community at the time when the theory predicting those phenomena was devised. There are various subversions of temporal predictivism. In some versions, temporally novel evidence is *generally* considered superior to accommodations. In other versions this may only *sometimes* be the case, in particular contexts. One may refer to these varieties as global and local forms of temporal predictivism (Hitchcock and Sober 2004). In yet other versions, temporally novel success can be a sign of another attractive, but opaque property of the theory, such as simplicity (Lipton 1991/2004, Hitchcock and Sober 2004, Harker 2008). One may refer to this brand of temporal predictivism as *symptomatic*. This may be contrasted with *intrinsic* forms of temporal predictivism where predicted evidence is somehow intrinsically better than accommodated evidence.⁶ At the core of all forms of temporal predictivism, however, stands the view that (at least sometimes) there is an epistemic asymmetry between temporally predicted evidence and already known evidence accommodated by a theory, to the effect that the former result in a higher degree of confirmation of the theory than the latter. I want to refer to this view as the *epistemic asymmetry thesis* of predictivism.

As mentioned in the introduction, Worrall has ferociously attacked temporal predictivism on the basis of historical case studies. Worrall’s work predates the various forms of temporal predictivism which have been distinguished, and strictly speaking, the case studies are compatible with local and symptomatic brands of temporal predictivism. But his studies still challenge the core of all views of temporal predictivism, namely the epistemic asymmetry thesis. One way of viewing Worrall’s case studies is that if it were to matter *at all* whether a phenomenon was

⁵ Other important accounts are comparative predictivism (Musgrave 1974), endorsement predictivism (Barnes 2008), views that deny any form of predictivism (Brush 1989, Mayo 1991, Harker 2008), and views that argue that accommodation is actually better confirmation (Dellsén forthcoming). See also the influential contributions by Lange (2001) and White (2003). For recent overviews see Barnes (2018) and Schindler (2018a).

⁶ Usually philosophers have used the terms “strong” vs. “weak” to describe “intrinsic” vs. “symptomatic” forms of predictivism. See e.g., Barnes (2018), who applies this distinction to all main predictivist accounts. I do not adopt this terminology here because I think intrinsic vs symptomatic is more specific and informative.

discovered after the theory was devised, then it surely should show in the historical cases of e.g., Fresnel and Mendeleev where we are faced with as astonishing predictive success as one can possibly imagine in science.

In a string of papers, Worrall has sought to replace temporal predictivism with his “use-novel account” or “heuristic predictivism” (Worrall 1989, Scerri and Worrall 2001, Worrall 2002; 2005).⁷ There are two versions of Worrall’s account, which actually incompatible, but which have not been distinguished in the literature (Schindler 2014). According to the more widely received version of heuristic predictivism a phenomenon is use-novel, when it was not used in the construction of the theory that predicts / entails it. According to another version, use-novelty is to be understood in terms of *free, adjustable parameters*, so that a piece of evidence E is use-novel for a theory when there are no free parameters in T that can be fixed to accommodate E. The two versions are incompatible because on the adjustable parameter version it does not seem to matter how the theory was constructed: it either does or does not possess free parameters for the evidence in question. Still, the two versions are presented by Worrall as just one coherent account (Worrall 2002; 2014), and it has been perceived as such by my most commentators.

Worrall believes that his account entails that there is no difference for Worrall confirmation-wise between use-novel accommodations and evidence that was temporally predicted. I call this view Worrall’s *epistemic symmetry thesis*. As we shall see in a moment, the epistemic symmetry thesis in fact only holds on the parameter-fixing version of his account.

Worrall also believes that the epistemic symmetry thesis (and therefore his account) is supported by available historical case studies, most notably by the Fresnel white spot prediction (Worrall 1989). For Worrall, though, the *known* straightedge diffraction patterns were clearly accommodated by Fresnel’s theory provided good support for the theory, because they weren’t used in the construction of the theory. Accordingly, there was no confirmatory asymmetry between the straightedge diffraction patterns and the white spot prediction.⁸ Worrall makes similar points about Mendeleev’s periodic table (Scerri and Worrall 2001).

A central motivation for Worrall’s use-novelty account is that theories should not accommodate the phenomena in an ad hoc fashion. As Worrall once put it, “it is wrong to regard the downgrading of ad hoc explanations and the apparent upgrading of genuine predictions as two separate methodological phenomena – they are at root the same phenomena” (Worrall 1989, 71). Phenomena that are used (ad hoc) in the construction of a theory do not constitute a genuine

⁷ Worrall’s account has been one of the most influential accounts of predictivism: it has for example been widely embraced in the scientific realism debate (Psillos 1999, Vickers 2013) and it has probably been the account most widely discussed and critiqued (Musgrave 1974, Mayo 1991, Barnes 2008, Harker 2008, Steele and Werndl 2013, Schindler 2014; 2018a).

⁸ Schindler (2014; 2018a) has argued that the historical facts, even by Worrall’s description, actually seem to indicate that the straightedge diffraction patterns were given *more* confirmatory weight than the white spot prediction (82).

test for the theory and therefore cannot lend any support to the theory (or if any, only very little). As Worrall explains:

if some particular feature of T was in fact tied down on the basis of e , so that T had been engineered to entail e , then checking e clearly constitutes no real test of T . It is obviously no test of T to ask it to get right some result which had been explicitly incorporated into it in the first place. (Worrall 1989, 148-9)

Indeed already Popper made the methodological connection between novelty and ad hoc-ness: writing about the requirement that theories ought to produce novel success, he argued that this requirement was “indispensable”, because “without it our new theory might be ad hoc; for it is always possible to produce a theory to fit any given set of explicanda” (Popper 1963/1978, 241-2). However, whereas Popper believed that *temporally novel* success was required in order to guard against ad hoc maneuvers, Worrall is confident that his heuristic account is best suited to account for our “intuitive” sense of testability (Worrall 1985, 313).

Is Worrall's dismissal of temporal predictivism perhaps too quick? Suppose the heuristic account is correct and it is problematic (at least in the first of Worrall's two versions) to use a piece of evidence E in the construction of a theory T that then entails E . But should evidence which we *cannot possibly* use in the construction of T not impress us even more? And temporally novel evidence is of course precisely such evidence, because we do not yet possess it when constructing T . In other words, even by the lights of Worrall's own account, temporally novel evidence ought to count more than any accommodated evidence. Combined with various other issues pertaining to the opacity of the construction process (e.g., how sure can we be whether scientist X used evidence E in the construction of T ? how accessible is that information to other scientists when they assess T ?), this conclusion seems even more inevitable (Schindler 2018a).⁹ At least on one version of his account, Worrall thus *unwittingly* provides a compelling rationale for temporal predictivism, which I want to refer to as *Ad hoc-ness Avoidance*:

Temporally novel success is superior to accommodated evidence, because the former (but not the latter) guarantee that the evidence was not accounted for in an ad hoc fashion.

Note that the ad hoc-ness avoidance rationale of predictivism of course does not entail that accommodated evidence can never be non-ad hoc; it may very well be. With temporally predicted evidence there is just no way the evidence can be accommodated in a non-ad hoc fashion. On a final note, one could say, with the terminology introduced at the beginning of this section, that the ad hoc-ness avoidance rationale implies a certain *symptomatic* form of temporal predictivism: when a theory has temporally novel success, then this is a sure sign that the theory did not accommodate the evidence in an ad hoc fashion. In contrast, when the evidence is not temporally novel, this matter may be opaque. It is important to stress that the ad hoc-ness avoidance rationale, and the

⁹ See also Gardner (1982) for criticisms of the heuristic account along those lines.

epistemic asymmetry between temporally novel success and accommodative success that it implies (but which Worrall rejects), only follow from the version of Worrall's account that emphasizes the construction process of the theory. In contrast the parameter-fixing version of Worrall's, in which the construction process is irrelevant, does entail the epistemic symmetry Worrall is advocating.

It is noteworthy that many contributions to the debate have shown sympathy to the idea that ad hoc-ness is a relevant concern motivating predictivism. For example, Lipton's "fudging explanation" of temporal predictivism can be read as a version of the ad hoc-ness avoidance rationale (Lipton 1991/2004).¹⁰ For accommodated data, Lipton believes, there is a "motive to force a theory and auxiliaries" so that the theory fits the data at the expense of the 'goodness' of the explanation (170). As Lipton puts it in a central passage: in the case of accommodations "the scientist knows the answer she must get, and she does whatever it takes to get it ... In the case of predictions, there is no motive for fudging, since the scientist does not know the right answer in advance" (170).¹¹ Hitchcock and Sober (2004), who have developed a local, symptomatic form of temporal predictivism, sympathize with Lipton's fudging explanation and associate fudging with overfitting the data.¹² An earlier paper by Forster and Sober, which is based on very similar lessons drawn from the modeling literature, makes the link between overfitting and ad hoc-ness even more explicitly (Forster and Sober 1994).

3 Accounts of ad hoc-ness

As we just saw, predictivism can and has been motivated by the ad hoc-ness avoidance rationale. We also noted that Worrall's use novelty account – contrary to what Worrall has argued and contrary to his epistemic symmetry thesis – actually entails that temporally novel success ought to 'count more' than even non-ad hoc accommodative success. Overall, use-novelty is in fact a rather problematic way of spelling out the notion of ad hoc-ness and, accordingly, the ad hoc-ness avoidance rationale of predictivism (Schindler 2018a). As already mentioned, there is an inherent opacity to the way in which scientists construct their theories, which stands in conflict with the objective assessment of a theory. There is also no evidence that scientists would explicitly mention

¹⁰ For example, White (2003) reads Lipton that way. Strangely, Lipton explicitly rejects the idea that fudging has anything to do with ad hoc accommodations (Lipton 1991/2004, 167).

¹¹ Lipton recognizes that he does not have much to say about how fudging is to be understood. But in passing, he does mention briefly that a fudged theory "becomes more like an arbitrary conjunction, less like a unified theory" (171) and that fudging of theories negatively affects the theory's "theoretical virtues" (180). See Lange (2001) for extending on this point and Harker (2008) for a critical discussion.

¹² An anonymous referee for this journal pointed out to me that Hitchcock and Sober frame their account as one that embraces the heuristic conception of novelty. But I think this is a misdescription for the same reason that the parameter-fixing version of Worrall's account is indifferent to information concerning the construction process: a theory or model does or does not possess a free parameter for the evidence in question.

or seek information about whether they used E in the construction of T when they want to accommodate E. And what exactly does it mean to use E in the construction of T in the first place?

There are several other alternative accounts of ad hoc-ness, which I want to briefly introduce in what follows.¹³ Perhaps somewhat confusingly, some of them entail a form of epistemic symmetry between temporally predicted evidence and certain kinds of accommodated evidence, and thus stand in tension with the epistemic asymmetry thesis of temporal predictivism (see section 2).

3.1 Parameter fixing

As mentioned in the previous section, Worrall's account comes in two versions. The parameter fixing version can be motivated in a similar way as the heuristic version: when E is used to fix a parameter in T, then E is no real test of T, because E is basically "written into" the theory. Accordingly, one may thus think that scientists may, for the sake of increasing the scientific standing of their theories, strive to construct theories that determine the central parameters. Conversely, a theory that contains too many parameters may be considered to be gerrymandered or ad hoc. Friederich et al. (2014) in fact suggest that the Higgs mechanism possessing many free parameters underlay physicists' judgments of it being ad hoc (prior to the discovery of the Higgs boson). As already mentioned in the previous section, Hitchcock and Sober (2004), and, in particular, Forster and Sober (1994), too may be thought of as entertaining the parameter fixing idea of ad hoc-ness.

It is quite obvious that on the parameter-fixing account, whether a piece of evidence E is temporally novel does not matter: as long as there is no free parameter for E, T's entailing of E would never be ad hoc. The parameter-fixing account thus seems incompatible with the ad hoc-ness avoidance rationale of predictivism.

3.2 Independent testability

The earliest attempt to explicate the concept of ad hoc-ness probably goes back to Popper. For him, ad hoc hypotheses are hypotheses that "cannot be tested independently" (Popper 1976, 986). In other words, ad hoc hypotheses cannot be tested by evidence other than one they were introduced to save. When introduced to save a theory, ad hoc hypotheses would reduce the theory's falsifiability, and therefore "degrees of ad hoc-ness are related (inversely) to degrees of testability and significance" (Popper 1959). This idea was challenged by Grünbaum (1959), who pointed out that Popper's prime example of an ad hoc hypothesis, namely the Lorentz-FitzGerald contraction

¹³ One may object that ad hoc hypothesis is simply one that is introduced to save a theory from refutation. For example, the Lorentz-FitzGerald contraction hypothesis – one of the most paradigmatic ad hoc hypotheses – was introduced to save the ether theories of the time from their apparent refutation through the notorious Michelson-Morley ether-null result experiment (see e.g., Leplin 1975). It has been pointed out, however, that this intuitive notion only tells us about motivations, not about what is epistemically or methodologically defective about ad hoc hypotheses (Barnes 2018, Schindler 2018a).

hypothesis, actually did have other empirical consequences than just accounting for the Morley-Michelson experiment (but see Holton (1969)).¹⁴ Since then, few philosophers have defended this notion of ad hoc-ness (Acuña 2014).

3.3 Independent support

Grünbaum (1976) himself concluded that ad hoc hypotheses are hypotheses which may make independent predictions, but for which there no support. This view is incredibly popular amongst philosophers (Schaffner 1974, Leplin 1975, Scerri and Worrall 2001, Worrall 2002, Sober 2008). The following example is often invoked to motivate the view (e.g., Worrall 2002): when an irregularity was discovered in the planet Uranus, Adams and Le Verrier in 1845-46 proposed that a new planet in the vicinity of Uranus might explain the discrepancy. At first, the “Neptune hypothesis” was ad hoc, because it was introduced to save Newton’s theory. But after Neptune was discovered, the hypothesis was independently confirmed and thus lost its ad hoc status.¹⁵ More recently, Friederich et al. (2014) have argued analogously that the Higgs mechanism in the standard model in fundamental particle physics lost its status as an ad hoc hypothesis, once the Higgs particle was discovered in 2013.¹⁶

3.4 Coherence

In previous work, I have myself suggested that an ad hoc hypothesis is one which does not cohere with the theory which it is invoked to save nor with the current background knowledge, whereby coherence is to be understood as good (i.e., empirically supported) theoretical reasons for belief (Schindler 2018a; b). For example, the accommodation of the phenomenon of maximum elongation of the inner planets¹⁷ by the Ptolemaic theory of planetary motion is ad hoc, because the theory (or any background theory) provides no good theoretical reason for belief that this should be so. Instead, Ptolemy’s theory arbitrarily stipulates, on the basis of the observations, that the center of an inner planet’s orbit had to be aligned with the center of the sun and the center of the earth. In the Copernican system, in contrast, maximum elongation obtains because the orbits of the inner planets are embedded in earth’s orbit – relative to the earth – so there is a physical limit to how far away the inner planets can be from the sun. I have also previously argued that other accounts of ad hoc-ness, such as parameter fixing, can be incorporated into his account (ibid.). For example, when there is a free parameter that needs to be fixed on the basis of the evidence, then the theory

¹⁴ Another empirical consequence of the contraction hypothesis can be tested in the so-called Kennedy–Thorndike experiment, which showed that the speed of light is independent of the velocity of the apparatus (the Michelson-Morley experiment shows the independence of the orientation of the apparatus). See also Janssen (2002).

¹⁵ It is interesting to note that there is no evidence that astronomers ever considered the Neptune hypothesis ad hoc, even before Neptune was discovered (Leplin 1982). It is thus questionable whether the case can be used by philosophers to learn anything about ad hoc hypotheses (Schindler 2018a).

¹⁶ See Schindler (2018a) for a critique of this claim.

¹⁷ Maximum elongation of the inner planets just means that the Mercury and Venus never can be seen removed from the ecliptic in the night sky beyond a certain angle.

(or our background knowledge) provides no theoretical reasons to believe that the parameter in question should have one or the other value.

4 An empirical study

This section presents the hypotheses, experimental design, and the results of the study conducted to probe scientists' views on predictivism and ad hoc-ness.

4.1 Hypotheses

The following hypotheses guided and structured the design of the study to be presented in the following sections. The hypotheses are not to be understood as endorsements.

H1: Evidence temporally predicted by a theory has higher confirmational value than already known evidence accommodated by the same theory.

H2: Ad hoc accommodations of the evidence should be avoided.

Hypothesis H1 targets the epistemic asymmetry thesis and asks whether there is a general preference among subjects for temporally novel success over accommodative success. Hypothesis H2 targets “ad hoc-ness avoidance” as perhaps the most general rationale for predictivism of any kind. Finally, the study explored what views scientists hold with regards to the nature of ad hoc-ness, and the relationship between H2 and scientists’ views of ad hoc-ness, without testing any specific hypotheses tested.

4.2 Participants

There were three groups of participants: natural scientists (214), social scientists (278), and philosophers of science (86).¹⁸ Subjects were recruited to email lists, social media, and via email to department heads. Participation requirement was a completed PhD in a relevant discipline or enrollment in a PhD program. The online platform *Qualtrics* was used to implement the survey. Participation was incentivized by an optional entry for an amazon voucher lottery. The average age of subjects were 38.6 in the group of natural scientists, 47.8 amongst social scientists, and 49.2 amongst philosophers. The percentage male was 80% among natural scientists, 74% among social scientists, and 71% among philosophers. Gender balance was not aimed for, as all of the three fields are unfortunately men-dominated at that level of professionalization. The average duration participants spent on the survey was 782 seconds (ca. 13 min). See also the appendix.

4.3 Study design

The survey consisted of three main blocks of questions with the headings: ‘theories’, ‘simplicity and unification’, and ‘intuitions’. There were also background questions before and after these three blocks. There are altogether thirteen topical questions and seven background questions. The three questions that were used for the current study were questions from the first block of topical

¹⁸ There were many more incomplete surveys, namely 226.

questions. Subjects were presented with at least one statement per question, which they had to take a stance toward. Per question, the statements were randomized. A pilot study was conducted prior to the conduction of the main study in order to ensure that subjects understood the questions in the way they were intended.¹⁹

4.4 Materials: study questions

The first question (relating to hypothesis H1) used for the current survey were a multiple-choice question on the conformational status of predicted versus accommodated evidence:

Q1: What do you think is better evidence for a scientific theory?

- *Known phenomena explained by a theory.*
- *Neither predicted nor known phenomena are better evidence; they count the same.*
- *New phenomena correctly predicted by a theory.*

Subjects could choose only one of these options.

A couple of remarks about Q1 are in order. First, note that Q1 uses the term ‘explanation’ instead of ‘accommodation’, which is more common in the predictivism literature. The reason is that the pilot of this study showed that using accommodation instead of explanation confused subjects: they apparently found the former to jargony. Second, the target of question Q1 obviously is the epistemic asymmetry thesis of temporal predictivism: subjects were given a choice between preferring the explanation of *known* phenomena or the prediction of *new* phenomena (or neither of these two options). The first choice (‘known phenomena explained by a theory’) and second choice (‘neither predicted nor known phenomena’) correspond to views that reject the epistemic asymmetry thesis of predictivism. Third, the third answer choice (‘new phenomena correctly predicted by a theory’) clearly would be a choice for the epistemic asymmetry thesis. Because the question does not specify any particular context, this answer choice could be read as a general preference, supporting global temporal predictivism. But subjects could also have local and symptomatic versions of temporal predictivism in mind.²⁰

¹⁹ Other results of this survey are published in Schindler (2022).

²⁰ An anonymous referee for this journal suggested to me that subjects may actually have situations in mind that undermine the significance of the results. For example, subjects may be thinking of situations in which the theory’s empirical consequences are severely tested by temporally novel evidence, but not severely tested by already known evidence. Their responses may thus express simply a preference for test severity (e.g., along the lines formulated by (Mayo 1991)) rather than for the asymmetry thesis of temporal predictivism. Although I agree that one cannot rule out anything when it comes to what subjects may be thinking, there is very little support for an interpretation suggested by the referee. After all, the contrast highlighted in the question is between “new phenomena” and “known phenomena” and *not* between “new, severely tested, phenomena” and “known, not severely tested, phenomena”. It would assume too much, I think, to assume that subjects have a notion of severity that is independent from temporal novelty and that subjects *just happen to* assign severity to the new phenomena but not to already known phenomena in the

Finally, we should also mention that Q1 does not specify in what way the evidence is accommodated, that is, whether it was accommodated ad hoc or not ad hoc. This was a conscious decision, so as to leave room for views that consider accommodated phenomena less confirmatory because there is a *risk* or *possibility* for ad hoc accommodations which there is not in predicted success, and it may not always be apparent whether or not accommodations are actually ad hoc fudged (Lipton 1991/2004). The phrasing of Q1 is also consistent with the ad hoc-ness avoidance rationale introduced in section 2, which says that predictive success is superior because it cannot possibly be ad hoc (whereas accommodations may be).

The second question (relating to hypothesis H2) was also a multiple-choice question concerning the use of ad hoc hypotheses:

Q2: *What do you think about the use of ad hoc hypotheses?*

- *Ad hoc hypotheses are nothing to worry about.*
- *Ad hoc hypotheses can be used under certain circumstances. Please state under which circumstances: _____.*
- *Ad hoc hypotheses should never be used in theorizing.*

The final, *exploratory* question (relating to no hypothesis) was a 5-point Likert scale question (ranging from 5=strongly agree to 1=strongly disagree) concerning the nature of ad hoc hypotheses:

Q3: *Ad hoc hypotheses are hypotheses that are introduced to save a theory from refutation. By what other features can you identify an ad hoc hypothesis? The hypothesis ...*

- *... has many free parameters.*
- *... has no independent empirical support.*
- *... is not testable.*
- *... doesn't cohere well with what we know.*

Note that a conscious decision was made not to include Worrall's use-novelty notion in the list of question. This has several reasons. First, there was concern that subjects would not understand what is meant by the use of evidence E in the construction of T. There is after all a sense in which any good theory ought to accommodate the evidence that it is facing (Worrall 2002). Second, and more importantly, part of what use-novelty means for Worrall, namely epistemic symmetry between temporally predicted and accommodated evidence, is already captured by the second statement of Q1 (see also section 2). Third, as mentioned in section 2, use-novelty is a highly problematic notion: it just seems highly implausible that the theory construction process itself, which is opaque to the community, ought to affect the community's assessment of the theory in

prompt. I see no reason why subjects would want to do that, unless they equated severity with temporal novelty (contrary to Mayo).

question. Worrall in fact accepts these arguments and argues that they do not undermine his account (Worrall 1985; 1989). But this is only true if use-novelty is read in terms of parameter-fixing (see Section 2). This is what the first sentence of Q3 tried to get at.

4.5 Results

The results that were obtained with regard to question Q1 ('what is better evidence') are summarized in Figure 1.

More than half of the natural scientists surveyed thought that new phenomena correctly predicted by a theory count more than known phenomena explained by a theory (54.7%). Philosophers agreed at exactly the same level (54.7%). About a third of each of these two groups agreed with the view that temporally predicted and accommodated evidence counts the same (33.6% and 32.6%, respectively). The fewest natural scientists and philosophers agreed with the view that "known phenomena explained by a theory" are better evidence than temporally predicted phenomena (11.7% and 12.8%, respectively). Thus, altogether 45.3% of natural scientists did not embrace temporal predictivism, versus 54.7% who did. Likewise, 45.4% of philosophers did not embrace temporal predictivism, versus 54.7% who did. In other words, a slight majority of natural scientists and philosophers came out in favor of H1.

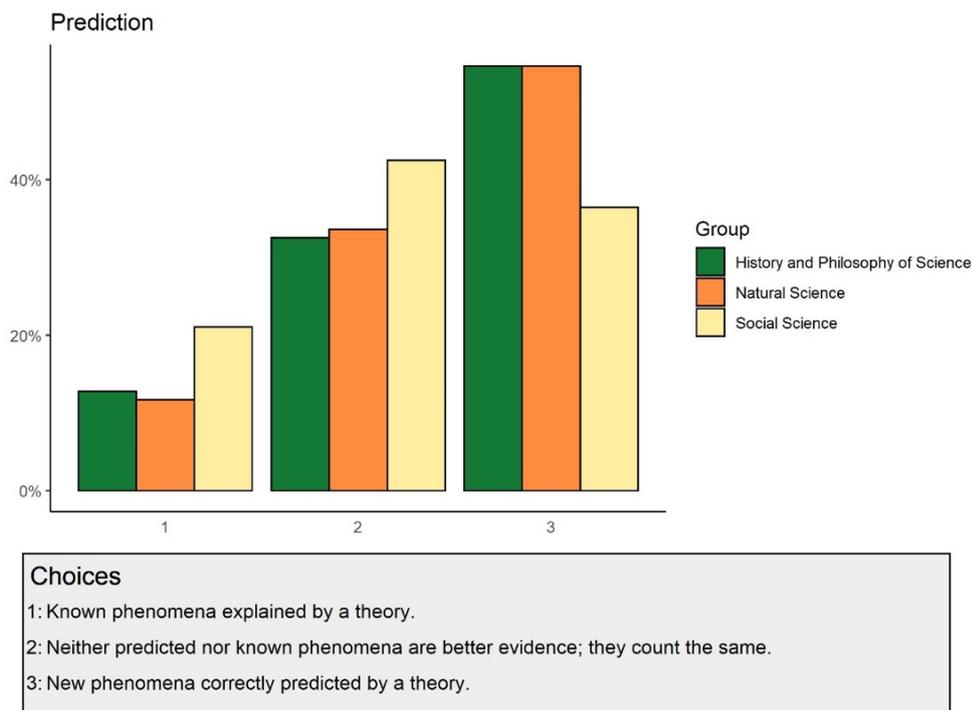


Figure 1: Diagram depicting the results for Q1. Agreement in percentage.

The results were quite different with regards to social scientists. Although more than a third of social scientists agreed with the view that correctly predicted phenomena count more than accommodations (36.4%), even more social scientists agreed with the view that predicted and accommodated evidence 'counts the same' (42.5%). There were also roughly twice as many social

scientists than natural scientists and philosophers who agreed that known phenomena explained by the theory count more than predicted phenomena (21.1%). Thus, altogether 63.6% of social scientists did not embrace temporal predictivism versus 36.4% who did. H1 is therefore disconfirmed with regards to social scientists.

The results concerning question Q2 ('the use of ad hoc hypotheses') are summarized in Figure 2. The agreement of all three groups was by far largest with the statement "ad hoc hypotheses can be used under certain circumstances": 54.2% of natural scientists, 62.9% of social scientists, and 61.6% of philosophers. About a quarter of natural and social scientists believes that ad hoc hypotheses should never be used in theorizing (26.2% and 24.6%, respectively). About one fifth of philosophers agrees with this view (18.6%). The smallest agreement among natural and social scientists alike was with the view that ad hoc hypotheses is nothing to worry about (19.6% and 12.5%, respectively). About one fifth of philosophers shared this view (19.8%). These results suggest that hypothesis H2 should therefore be rejected (with respect to all three groups): ad hoc hypotheses *can* be used, but only under certain circumstances.

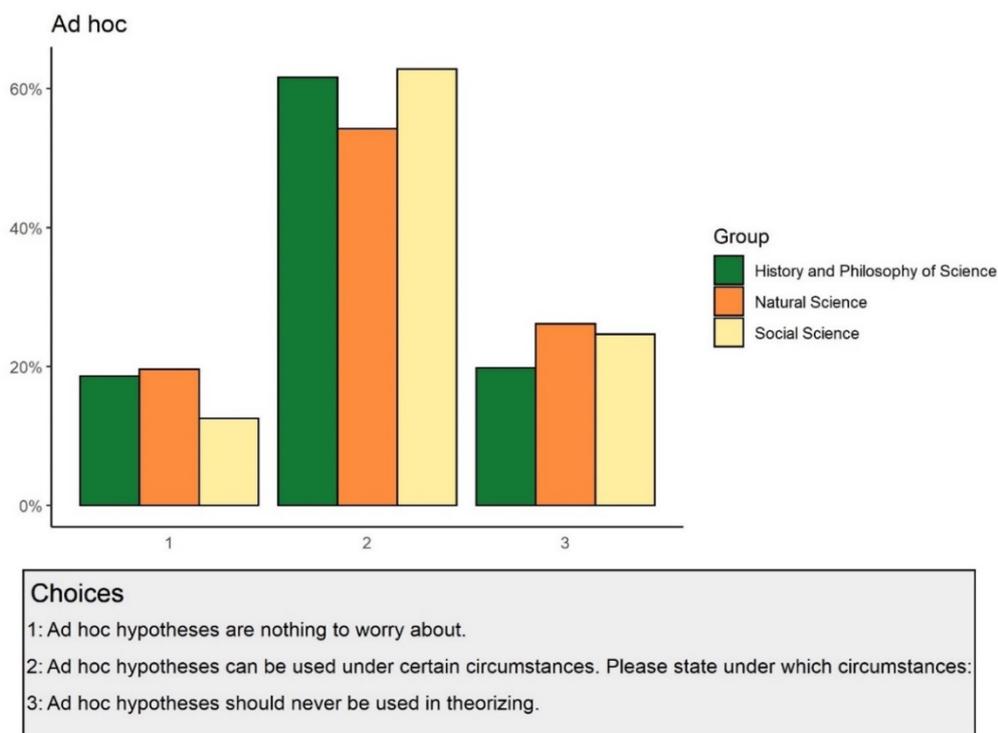


Figure 2: Diagram depicting the results for Q2. Results in percentage.

Of the subjects who opted for the statement that "ad hoc hypotheses can be used under certain circumstances", altogether 281 stated under which circumstances they thought that would be the case (NAT: 97 subjects, SOC: 137 subjects, HPS: 47 subjects). All responses were manually coded and categorized. Table 1 shows which categories occurred most frequently.^{a21}

²¹ Some responses were assigned more than one category.

	NAT	SOC	HPS
“when they help to develop a new theory”	17	36	
“when the hypothesis is testable”	16	18	
“when there are no alternatives”	14	13	8
“acceptable only temporarily”	11	14	5
“when one is explicit about the hypothesis being ad hoc”	8	8	
“when the issue is relatively minor”			5
“when there is independent support”			5

Table 1: Manually coded categories of the most frequently mentioned conditions under response 2 to Q2.

It is quite apparent from these responses that, despite scientists (and philosophers) rejecting the notion that ad hoc hypotheses are categorically bad (as suggested by H2), it still is the case that scientists (and philosophers) view ad hoc hypotheses as not at all ideal. The results of this study therefore conflict with the view that ad hoc judgements are merely aesthetical and subjective (see Hunt 2012).

The results concerning question Q3 (5-point Likert scale) are depicted in Figure 3. Natural scientists and social scientists agree most strongly with the view that ad hoc hypotheses have too many free parameters (NAT: mean=3.53, SD=1.02; SOC: mean=3.65, SD=0.96) and that ad hoc hypotheses don't have independent empirical support (NAT: 3.68, SD=1.08; SOC: 3.64, SD=1.11). Both groups found the other two views less plausible: natural scientists and social scientists only slightly agree that ad hoc hypotheses are not testable (NAT: 3.38, SD=1.36; SOC: 3.37, SD=1.33) or that they lack coherence (NAT: 3.36, SD=1.08; SOC: 3.23, SD=1.08). Philosophers had a clear preference for the view that ad hoc hypotheses don't have independent empirical support (3.8, SD=1.23). Philosophers also had some sympathy for the view that ad hoc hypotheses have too many free parameters (3.37, SD=1.02), but were not enthused by the view that ad hoc hypotheses are not testable (3.13*, SD=1.33) or that ad hoc hypotheses lack coherence (2.97*, SD=1.21).²² There were no criteria for which any group showed on-average disagreement. See also the appendix.

²² The starred average values indicate that there is no significant difference from the midpoint of the 5-point scale.

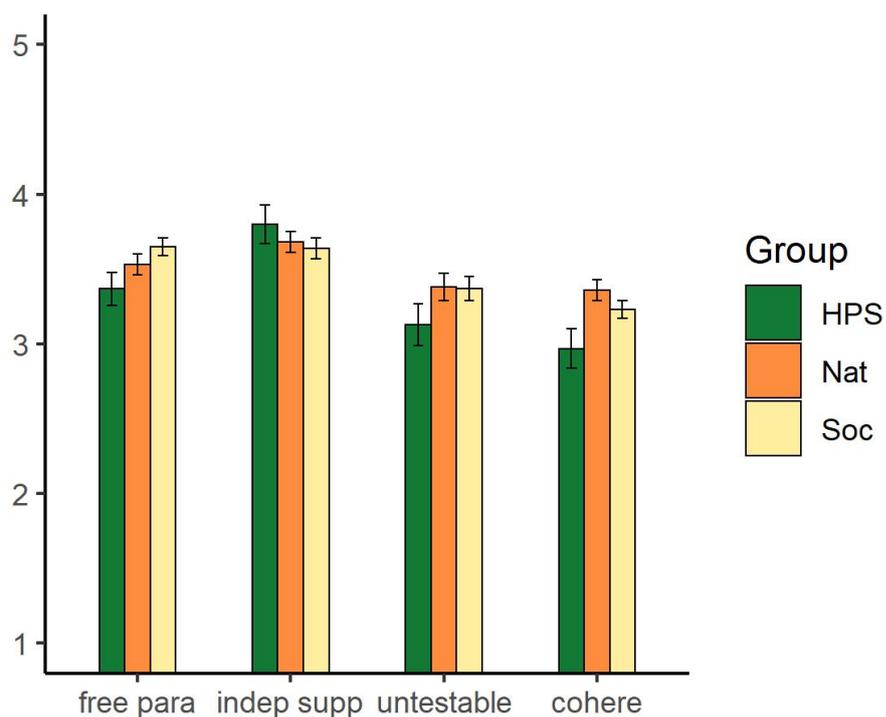


Figure 3: Diagram depicting the results for question Q3, concerning the nature of ad hoc hypotheses. Error bars indicate standard errors.

An ANOVA analysis revealed no statistically significant differences between the three groups for any of the four criteria of ad hoc-ness, except for coherence: $F(2.577)=3.972$, $p=0.019$. A pairwise, post hoc Tukey test showed that only the difference between the natural science group and the HPS group was significant ($p=0.014$). A further ANOVA analysis revealed that the differences in the responses for each of the four criteria were statistically significant within all of the groups: NAT: $F(3,852)=3.73$, $p<0.01$; SOC: $F(3,1116)=9.73$, $p<0.01$; HPS: $F(3,340)=7.88$, $p<0.01$. Post hoc Tukey tests determined that there were the following significant differences: NAT: independent support > not testable & not coherent ($p<0.05$, respectively); SOC: independent support > not coherent ($p<0.01$) & not testable ($p<0.05$); free parameters > not coherent ($p<0.01$) and not testable ($p<0.05$); HPS: independent support > not coherent & not testable ($p<0.01$, respectively).

Since the midpoint of Likert scales is notoriously problematic (Nadler et al. 2015), the results were also analyzed without the midpoint results. Merging the two upper and lower ends of the scale, the highest level of agreement among all three groups was found for the view that ad hoc hypotheses don't have independent empirical support: NAT: 60.3% vs. 13.1% disagree; SOC: 59.6% vs. 18.2% disagree; HPS: 67.5% vs. 15.1% disagree. The next most popular view was that ad hoc hypotheses have too many free parameters: NAT: 54.2% vs. 12.6% disagree; SOC: 59.3% vs. 10% disagree, HPS: 51.2% vs. 19.8% disagree. All groups had the highest level of disagreement for the view that ad hoc hypotheses are not testable: NAT: 50.9% agree vs. 30.4% disagree; SOC: 48.2% agree vs. 28.6% disagree; HPS: 44.2% agree vs. 37.2% disagree. The view that ad hoc hypotheses

are those that lack coherence had less disagreement, but also comparatively low levels of agreement: NAT: 43.5 agree vs. 18.7% disagree; SOC: 40% agree vs. 26.1% disagree; HPS: 39.5% agree vs. 36.1% disagree. See the appendix for more details. Note that this analysis, which excludes the midpoint of the scale, is consistent with the results revealed by the ANOVA and post-hoc Tuckey tests.

Finally, the study probed whether there were any correlations between how subjects answered Q2, namely positively (“nothing to worry about”), negatively (should never be used”), or neutrally (“can be used under certain circumstances”), and their answers to Q3 about the nature of ad hoc-ness. As can be seen from figure 4, there are visible differences for ad hoc-ness as untestable: subjects who had a negative view of the use of ad hoc hypotheses were quite likely to strongly agree with this view of ad hoc-ness, and those who had a more positive view were more likely to disagree with this view of ad hoc-ness. Something similar (but less prominent) is the case for the independent support notion of ad hoc-ness, where a negative view of the use of ad hoc hypotheses mostly affected the agreement with this notion, but where subjects with a positive view were not more likely to disagree with this notion. The other two forms of ad hoc-ness, namely parameter freedom and coherence did not show any such pattern.

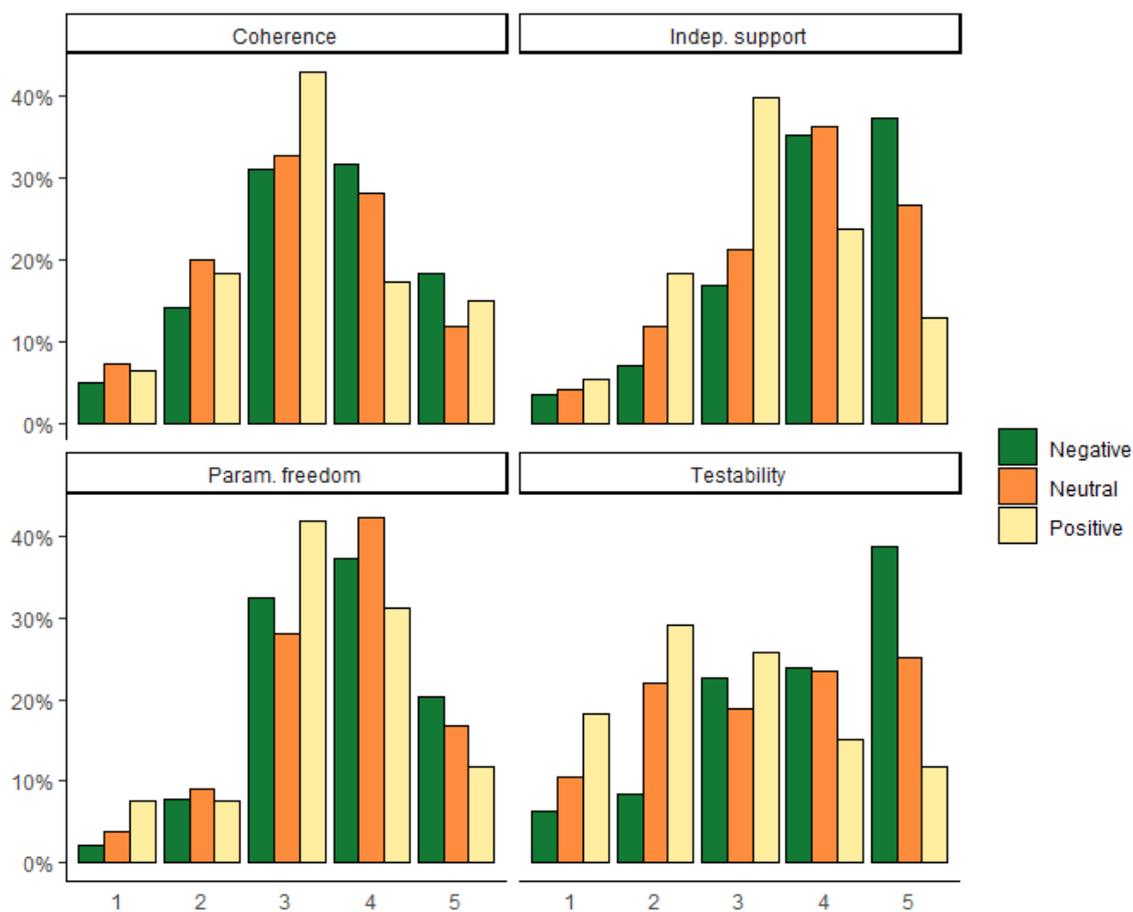


Figure 4: Subjects' responses to Q3, concerning the nature of ad hoc hypotheses, divided by negative, neutral, and positive responses to Q2 about whether the use of ad hoc hypotheses is permissible.

5 Discussion and conclusion

Previous historical evidence has indicated that scientists do not favor highly impressive temporally novel predictions over accommodations of already known phenomena. In the face of the historical facts, some philosophers have therefore rejected what I called the epistemic asymmetry thesis of temporal predictivism. Using historical case studies to empirically ground the predictivism debate, however, is not unproblematic. For example, the historical evidence is only circumstantial and tends to focus on the epistemic views of individual scientists. It was therefore one of the main goals of the paper to obtain data that provides more *direct* evidence of the *community* view by conducting an empirical study. The other main goal the paper pursued was to probe empirically what I called the *ad hoc-ness rationale of predictivism* that *prima facie* seems to justify the epistemic asymmetry thesis.

Contrary to what the historical facts have suggested, the results of the present study indicate that the responses of a small majority of natural scientists actually is in line with the epistemic asymmetry thesis of temporal predictivism. A similar result was obtained for philosophers. Although still a third of social scientists agrees that novel phenomena predicted by a theory count more than accommodated phenomena, the majority of social scientists seems to reject the epistemic asymmetry thesis.

The results raise questions about disciplinary differences that have not yet been captured at all in the philosophical debate about predictivism. Usually, it is presumed that the right answer to the question of whether predictive success counts more than accommodative success should have the same answer for *all* sciences. But that assumption looks quite mistaken. This is not something that the historical case studies have shown, or could have shown for that matter, since they were all drawn from the natural sciences. So far, none of the extant accounts has sought to reflect such differences between disciplines. On the contrary, philosophers have hitherto offered accounts that were meant as one-that-fits-all accounts. In fact, there may even be more fine-grained differences between different academic disciplines within the two very broad fields of social science and natural science, which this study was not powerful enough to detect. Future studies of the question may attempt to improve on the current study on this issue, although pragmatic constraints, such as getting a sufficient number of subjects for each discipline, may make this difficult.

Why do social scientists not favor temporally novel evidence? There is an optimistic and a pessimistic view. On the pessimistic view, social scientists should value temporal predictions more than they do. If social scientists were only more eager to generate (falsifiable) predictions, then their field might be in a better state than it actually is (think replication crisis, for example). On the optimistic view, the phenomena social science deals with tend to be very complex and the contributing causal factors hard to disentangle; experiments isolating those factors are often hard to come by or even be unethical. All of that might entail that reliable predictions about the

evolution of social systems and phenomena may be hard to make. It is for this reason, so the thought might go, that predictive success plays only a minor role in the social sciences. Whether the optimistic or the pessimistic view (or a third view) actually holds may have to be determined by future empirical studies.

As to the rationale of predictivism, this study explored only the most general rationale that has been offered in the literature, namely a rationale that I called “ad hoc-ness avoidance”. The study showed that only a minority of scientists agreed with a categorical rejection of the claim that ad hoc hypotheses should never be used in theorizing. Another minority agreed with the view that they are “nothing to worry about”. The majority of scientists turned out to have more nuanced views about the use of ad hoc-ness: they think that the use of ad hoc hypotheses is permissible under certain circumstances. The circumstances that were mentioned most frequently were circumstances in which no alternatives were available to ad hoc hypotheses, circumstances in which ad hoc hypotheses can help develop a new theory, when ad hoc hypotheses are used only temporarily, and when they are explicitly flagged as ad hoc hypotheses. Philosophers’ views are well-aligned with all of this. These results indicate that even though the use of ad hoc hypotheses is not deemed outright impermissible by scientists (and philosophers), scientists (and philosophers) still view the use of ad hoc hypotheses as not ideal and as something to be avoided if possible (contrary to the subjectivist view e.g., defended by Hunt (2012)). This view, I think, is indeed broadly compatible with the ad hoc-ness avoidance rationale of predictivism.

As to the question of what it is that renders ad hoc hypotheses epistemically problematic, the results of this study showed that social and natural scientists have a preference for the view that ad hoc hypotheses have no independent empirical support. Philosophers agree. The view that ad hoc hypotheses have too many free parameters is the second most popular view among all three groups, intriguingly especially among social scientists. Both views have also been popular in the philosophical literature, both on the topic of predictivism and on the topic of ad hoc-ness. The view that ad hoc hypotheses are not testable had the least adherence among all three groups. Still all four views (including the view that ad hoc hypotheses have to do with a lack of coherence) on average enjoy positive support among scientists: there was no view for which the mean response indicated disagreement.

Finally, the study found correlations between subjects’ rejection of the use of ad hoc hypotheses and their agreement with the notion that ad hoc hypotheses are untestable, and vice versa, between subjects’ endorsement of ad hoc hypotheses and their rejection of the view that ad hoc hypotheses are untestable. A weaker correlation was found also for subjects’ rejection of ad hoc hypotheses and the notion of independent support. It is not clear what would explain these correlations, but one possible explanation for the first (stronger) correlation is that subjects may have been influenced by Popper, who held that ad hoc hypotheses are not testable (or at least decrease the degree of testability), and whose demarcation criterion implied a strong rejection of the use of ad hoc hypotheses. However, this influence, if it does exist, does not seem very strong,

as this notion of ad hoc-ness was the least popular notion among the four views presented to the subjects.

The results of this study are perhaps the first data of their kind regarding the attitudes of scientists towards the epistemic asymmetry thesis of temporal predictivism and ad hoc-ness, and therefore should be interesting in and of themselves. Again, there is no longer any need to infer scientists' epistemic attitudes from circumstantial historical evidence, as it also has been the case hitherto both in the literature on predictivism and ad hoc-ness (see Schindler 2018a). That alone is progress. Beyond that, I think the results of the study can only serve as the basis for further discussion.

On the one hand, the fact that a majority of scientists came out embracing the epistemic asymmetry thesis and were in broad favor of the avoidance of ad hoc hypotheses seems to support both temporal predictivism and the ad hoc-ness avoidance rationale. This rationale, in turn, could be understood as a form of symptomatic predictivism: a theory's temporally novel success can be considered a sure sign that the theory did not accommodate the evidence in an ad hoc fashion, when normally this may not be so obvious. On the other hand, as we have seen, some of the ways in which the notion of ad hoc-ness has been spelled out by philosophers hitherto all seem to entail an epistemic symmetry between temporally predicted evidence and certain kinds of accommodated evidence. This of course is in tension with the ad hoc-ness avoidance rationale. Future research will have to resolve this conflict. There are several options of how to resolve it: (i) one could argue, in principle, that the survey results are not what they appear to be and stick to one's guns; (ii) one could reject the idea that ad hoc-ness avoidance is the best rationale of temporal predictivism to begin with and suggest a different rationale; or, (iii) one could propose a new account of ad hoc-ness that does not entail the epistemic asymmetry thesis. It will be to philosophers to figure out the most viable option.

I want to close by making a couple of methodological remarks about a study like the current one. First of all, one might outright reject the relevance of scientists' attitudes towards predictive success for the predictivism debate. Why, one might ask, should scientists' attitudes matter at all for philosophical disputes? As indicated already in the introduction of the present paper, philosophers have as a matter of fact *already* heeded scientists' attitudes in their discussions of predictivism, namely in the form of historical case studies. These case studies provided *circumstantial* evidence for such attitudes, for example in the form of the prize committee of the French academy paying less attention to Fresnel's predictive success than to his explanatory success. The current study provides more direct data about scientists' attitudes than the historical case studies have. Hence, if one thinks that historical case studies bear on the predictivism debate in one way or another (which most partakers of the debate have), then one cannot dismiss the results of a study like the current one as irrelevant. Even regardless of this conditional, I think one should have an interest for scientists' attitudes when thinking about science and predictivism in particular, because it is scientists who need to weigh the evidence in a particular way when

assessing their theories. I cannot think of a more relevant source of information when it comes to the predictivism debate.

Sometimes one hears (informally) that scientists may actually not be suited to answer philosophical questions about science and that scientists' views are simply too diverse for philosophers to build anything upon them. Both of these worries can be repudiated on the basis of the results themselves: if scientists had no opinion on the views offered to them in the study, one would have expected them to choose the midpoint of the scale (neither agree nor disagree) much more often than they did. And if scientists' opinions were as diverse as some philosophers sometimes intimate, then one wouldn't expect there to be discernable majority views (as there were in this study). There may of course be a self-selection issue here: the scientists who participated may be more interested than the average scientist in the type of questions posed in this study, so one cannot infer from this survey that scientists *in general* reliable informants about philosophical questions.

Another, related concern might be that even if scientists do have views on philosophical questions about their practice, these views may not necessarily reflect what the practice is actually like. In other words, scientists' words may not match their deeds. This possibility cannot be repudiated by the current study. But until it has systematically been proven that this possibility is real and systematic, I think we should give scientists the benefit of the doubt.

No empirical evidence can of course replace philosophical theorizing in philosophical discussions; the former can only be the beginning of the latter. All the more reason to make sure that the premises of our theorizing are as accurate as they can be.

References

- Acuña, Pablo. 2014. On the empirical equivalence between special relativity and Lorentz's ether theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **46, Part B** (0): 283-302.
- Barnes, Eric C. 2008. *The Paradox of Predictivism*. Cambridge: Cambridge University Press.
- . 2018. Prediction versus Accommodation. *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*, edited by Edward N. Zalta, <https://plato.stanford.edu/archives/fall2018/entries/prediction-accommodation>.
- Brush, Stephen G. 1989. Prediction and theory evaluation: The case of light bending. *Science*, **246** (4934): 1124-1129.
- Chang, Hasok. 2011. Beyond Case-Studies: History as Philosophy. In *Integrating History and Philosophy of Science*, Seymour Mauskopf and Tad Schmaltz (eds.), Heidelberg: Springer, 109-124.
- Dellsén, Finnur. forthcoming. An Epistemic Advantage of Accommodation over Prediction. *Philosophers' Imprint*.
- Forster, Malcolm and Elliott Sober. 1994. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, **45** (1): 1-35.

- Friederich, Simon, Robert V Harlander, and Koray Karaca. 2014. Philosophical perspectives on ad hoc hypotheses and the Higgs mechanism. *Synthese*:
- Gardner, Michael R. 1982. Predicting novel facts. *British Journal for the Philosophy of Science*, **33** (1): 1-15.
- Glymour, Clark. 1980. *Theory and evidence*. Princeton: Princeton University Press.
- Grünbaum, Adolf. 1959. The falsifiability of the Lorentz-Fitzgerald contraction hypothesis. *The British journal for the philosophy of science*, **10** (37): 48-50.
- . 1976. Ad hoc auxiliary hypotheses and falsificationism. *The British Journal for the Philosophy of Science*, **27** (4): 329-362.
- Harker, David. 2006. Accommodation and prediction: The case of the persistent head. *The British journal for the philosophy of science*, **57** (2): 309-321.
- . 2008. On the predilections for predictions. *The British journal for the philosophy of science*, **59** (3): 429.
- Hitchcock, Christopher and Elliott Sober. 2004. Prediction versus accommodation and the risk of overfitting. *The British journal for the philosophy of science*, **55** (1): 1-34.
- Holton, Gerald James. 1969. Einstein, Michelson, and the "crucial" experiment. *Isis*, **60** (2): 133-197.
- Hunt, J. Christopher. 2012. On Ad Hoc Hypotheses. *Philosophy of Science*, **79** (1): 1-14.
- Janssen, Michel. 2002. Reconsidering a scientific revolution: The case of Einstein versus Lorentz. *Physics in Perspective*, **4** (4): 421-446.
- Kinzel, Katherina. 2015. Narrative and evidence. How can case studies from the history of science support claims in the philosophy of science? *Studies in History and Philosophy of Science Part A*, **49**: 48-57.
- Lakatos, Imre. 1978. *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers*. Edited by John Worrall and Gregory Currie. Cambridge: Cambridge University Press.
- Lange, Marc. 2001. The apparent superiority of prediction to accommodation as a side effect: a reply to Maher. *The British journal for the philosophy of science*, **52** (3): 575-588.
- Leplin, Jarrett. 1975. The concept of an ad hoc hypothesis. *Studies In History and Philosophy of Science Part A*, **5** (4): 309-345.
- . 1982. The assessment of auxiliary hypotheses. *The British Journal for the Philosophy of Science*, **33** (3): 235-249.
- Lipton, Peter. 1991/2004. *Inference to the Best Explanation*. 2nd ed. London: Routledge.
- Maher, Patrick. "Prediction, accommodation, and the logic of discovery". *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Arthur Fine and Jarrett Leplin, Vol. 1988, Vol. One: Contributed Papers: 273-285.
- Mayo, Deborah G. 1991. Novel evidence and severe tests. *Philosophy of science*, **58** (4): 523-552.
- Musgrave, Alan. 1974. Logical versus historical theories of confirmation. *British Journal for the Philosophy of Science*, **25** (1): 1-23.
- Nadler, Joel T., Rebecca Weston, and Elora C. Voyles. 2015. Stuck in the Middle: The Use and Interpretation of Mid-Points in Items on Questionnaires. *The Journal of General Psychology*, **142** (2): 71-89.
- Popper, Karl R. 1959. Testability and 'ad-hocness' of the contraction hypothesis. *British Journal for the Philosophy of Science*, **10** (37): 50.
- . 1963/1978. *Conjectures and Refutations: The Growth of Scientific Knowledge*. fourth edition ed. London: Butler & Tanner Limited.
- Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.

- Saatsi, Juha. 2019. Historical Inductions, Old and New. *Synthese*, **196**: 3979–3993.
- Scerri, Eric R. and John Worrall. 2001. Prediction and the periodic table. *Studies in History and Philosophy of Science Part A*, **32** (3): 407-452.
- Schaffner, Kenneth F. 1974. Einstein versus Lorentz: research programmes and the logic of comparative theory evaluation. *The British Journal for the Philosophy of Science*, **25** (1): 45-78.
- Schindler, Samuel. 2014. Novelty, coherence, and Mendeleev's periodic table. *Studies in History and Philosophy of Science Part A*, **45** (March): 62-69.
- . 2018a. *Theoretical Virtues in Science: Uncovering Reality Through Theory*. Cambridge: Cambridge University Press.
- . 2018b. A coherentist conception of ad hoc hypotheses. *Studies in History and Philosophy of Science*, **67** (February): 54–6.
- . 2022. Theoretical Virtues: do scientists think what you think they ought to think? *Philosophy of Science*, **89** (3): 542-564.
- Sober, Elliott. 2008. *Evidence and evolution: the logic behind the science*. Cambridge: Cambridge University Press.
- Sprenger, Jan and Stephan Hartmann. 2019. *Bayesian philosophy of science*. Oxford: Oxford University Press.
- Steele, Katie and Charlotte Werndl. 2013. Climate models, calibration, and confirmation. *The British Journal for the Philosophy of Science*, **64** (3): 609-635.
- Vickers, Peter. 2013. A Confrontation of Convergent Realism. *Philosophy of Science*, **80** (2): 189-211.
- White, Roger. 2003. The epistemic advantage of prediction over accommodation. *Mind*, **112** (448): 653-683.
- Worrall, John. 1985. Scientific Discovery and Theory-Confirmation. In *Change and Progress in Modern Science*, Joseph C. Pitt (ed.), Dordrecht: D. Reidel, 301-331.
- . 1989. Fresnel, Poisson and the 'White Spot': The Role of Successful Prediction in Theory-acceptance. In *The Uses of Experiment*, David Gooding, Trevor Pinch and Simon Schaffer (eds.), Cambridge: Cambridge University Press, 135-158.
- . 2002. New evidence for old. In *In the Scope of Logic, Methodology and Philosophy of Science*, Peter Gardenfors (ed.), Dordrecht: Kluwer, 191-209.
- . 2005. Prediction and the 'periodic law': a rejoinder to Barnes. *Studies in History and Philosophy of Science Part A*, **36** (4): 817-826.
- . 2014. Prediction and accommodation revisited. *Studies in History and Philosophy of Science Part A*, **45**: 54-61.