Theoretical virtues: do scientists think what philosophers think they ought to think?

Abstract

Theoretical virtues play an important role in several debates in the philosophy of science (theory choice, realism, values, laws of nature). This paper presents the results of a quantitative survey that was conducted to investigate the views that natural and social scientists and HPS scholars hold concerning theoretical virtues. Some of the main results are: (i) there is broad agreement across all three groups about how the virtues are to be ranked in terms of importance, (ii) all groups agree that unification is an epistemic virtue, but results concerning simplicity are indeterminate, (iii) all three groups view syntactic parsimony as a form of simplicity, but are indifferent / undecided about ontological parsimony, and (iv) all groups find simplicity and unification aesthetically pleasing.

1 Introduction

Theoretical virtues – the good-making features of theories – are a central topic in the philosophy of science. Consistency, accuracy, simplicity, unifying power, and fertility. figure prominently in debates about scientific realism and theory choice (Kuhn 1977, van Fraassen 1980, McMullin 1982, Forster and Sober 1994, McMullin 1995, Okasha 2011, Douglas 2014, Schindler 2018) and have also been discussed as "explanatory virtues" in the rich literature on the Inference to the Best Explanation (Harman 1965, Barnes 1995, Lipton 2004, Schupbach 2011). The literature on value judgments in science, too, includes discussions of the role of theoretical virtues (Douglas 2009, Steel 2010). Finally, theoretical virtues have played a role in the literature on laws of nature, and in particular Lewis' best systems analysis (Lewis 1973, Woodward 2014).

When philosophers ponder whether particular virtues are desirable, truth-conducive, or whether there are any trade-offs between the virtues, philosophers have hitherto almost exclusively relied on historical case studies, studies of bits of contemporary scientific practice, and common sense.[1] However, alternative ways of grounding one's philosophical theorizing have emerged recently in the form of experimental methods.

Experimental methods have been employed in several subfields of philosophy since the early 2000s (Knobe and Nichols 2017). Although there have been early attempts to use experimental methods in the philosophy of science (Stotz and Griffiths 2004, Stotz et al. 2004, Stotz 2009), it has only been more recently that experimental methods are being embraced more widely (Schupbach 2011, Douven and Schupbach 2015, Schupbach 2017, Steel et al. 2017, Chall et al. 2019,

[1] Mizrahi (forthcoming) is an exception. It will be discussed later in the paper (Section 5).

Mättig and Stöltzner 2019, Robinson et al. 2019, Beebe and Dellsén 2020, Mizrahi 2020, forthcoming).[2] The current paper is a contribution to this recent wave in experimental philosophy of science that seeks to compare philosophical accounts to the views held by scientists.

There is reason to think that the views articulated by philosophers should be embraced by scientists. If philosophers were to articulate views about scientific practice that were unrecognizable to scientists, or if philosophers would even articulate views in conflict with how scientists think about their own practices, then scientists would be deceived about what they do. This is not impossible, but it would require a substantive explanation for why we should think that scientists are so deceived. Of course, the divergences between scientists' and philosophers' views may be less extreme and there may be instances in which philosophers do have better justifications for views discordant with scientists' views. The study sought to test whether there are such divergences.

The paper will proceeds as follows. Section 2 will provide motivations for the major hypotheses that the survey tested. Section 3 will present the method of the study, Section 4 the results, Section 5 will discuss weaknesses of the study and Section 6 will give a conclusion.

2 Motivations and major hypotheses

In one of the first philosophical publications on the topic of theoretical virtues, Kuhn (1977) gave what is now considered a standard list of theoretical virtues: consistency (internal and external), empirical accuracy, scope (or unifying power), simplicity, and fertility. Kuhn believed that the weighing of the virtues in theory choice was a rather subjective matter and that the diversity of preferences would make it so that there is no unique algorithm for theory-choice (see Okasha 2011).[3] But is that actually the case? This is an empirical question, which motivated the following hypothesis in this study:

H1: the majority of scientists prefer some theoretical virtues to others.

Simplicity, due to its inherent vagueness, has been criticized by many as particularly problematic in theory choice (Kuhn 1977, McAllister 1999, Achinstein 2018). In fact, it is probably fair to say that simplicity is the most criticized theoretical virtue in the philosophy of science (see Schindler 2018). If scientists had sympathies to such a view, one would expect them to rank simplicity comparatively low. Hence:

H2: scientists rank simplicity lower than other theoretical virtues.

A central focus of many philosophical discussions of theoretical virtues is the question of whether virtues like simplicity and unifying power are epistemic virtues, that is, whether they are truth-conducive. Famously, this question has been emphatically denied by the antirealist van Fraassen (1980, 1989), who argued that simplicity and unifying power in particular are merely *pragmatic* virtues, that is, virtues that concern the use of theories, but not their truth content. This

[2] See also Machery (2016) for an overview, which also includes other empirical methods in philosophy of science, such as anthropological methods.

[3] Okasha has argued for something stronger, namely that an algorithm for theory choice is impossible.

view has been embraced by many others, even outside the realism debate (Hacking 1982, Barnes 1995, Douglas 2009, 2014, Achinstein 2018, Wray 2018). We tested whether this view was shared by scientists:

H3: scientists view simplicity / unification as merely pragmatic virtues.

Even though simplicity is a vague notion, some broad forms of simplicity have been identified, namely syntactic parsimony (simplicity of principles) and ontological parsimony (paucity of postulated entities) (Baker 2016). Paucity of parameter freedom is also sometimes discussed as a form of simplicity (Forster and Sober 1994). The survey sought to find out whether scientists prefer any one of these forms.

H4: scientists prefer a particular notion of simplicity to others.

It is well known that scientists sometimes judge theories as aesthetically pleasing (McAllister 1999, Ivanova 2017, Ivanova and French 2020). It is however not clear how ubiquitous this sentiment is. It is also unclear what properties underlie judgments that theories are aesthetically pleasing. The survey sought to find out by testing the following hypothesis:

H5: scientists consider theories (with certain properties) aesthetically pleasing.

A more overarching hypothesis that this survey tested was whether the views articulated by scholars in HPS would actually reflect the views by practitioners working in the fields studied by HPS scholars.

H6: there is disagreement between scientists and scholars in HPS.

Before proceeding to the presentation of the study and the results, a couple of clarifications are in order with regard to Kuhn's list of standard theoretical virtues. First, it makes good sense to distinguish internal and external consistency (see Douglas 2014, Schindler 2018). Internal consistency is the absence of any contradictions within a theory, whereas external consistency is the absence of contradictions with other (empirically established) theories. Second, Kuhn did not have much to say about what he meant by the fertility of a theory.[4] There are different ways of interpreting fertility, but perhaps the most popular interpretation is that a theory is fertile when it has novel success (Schindler 2018).[5] It is this sense of fertility on which this paper will focus. Third, there are two senses of scope or unification. According to one conception, a unifying theory is just more empirically accurate than one that doesn't unify the relevant phenomena. This "deflationary" conception of unification, as one might call it, simply reduces to the virtue of accuracy. But there is also another conception of unification, namely unification as identifying principles underlying seemingly disparate phenomena (Schindler 2018). For example, Maxwell's theory identified light as a form of electromagnetic radiation, and in the 1970s high energy physicists unified electromagnetic and weak interactions in a precursor of the standard model (Morrison 2000). In order to keep unification distinct from accuracy, the latter was used in this survey.

[4] Kuhn (1977) described fertility is a theory's capacity to "disclose new phenomena or previously unnoted relationships among those already known" (322).

[5] For alternatives see McMullin (1976), Schindler (2017), and Ivani (2018).

3 Methods

3.1 Study design

A survey consisting of three main blocks (5 + 5 + 3 questions) and embracing background questions (2 + 5) was used. The questions of the first block related to the purpose of theories (Q1.1), the ranking of theoretical virtues (Q1.2), prediction and ad hocness (Q1.3-1.5), the epistemicity of virtues (Q2.2 and Q2.3), the nature of virtues (Q2.1, Q2.4), aesthetics and virtues (Q2.5) and intuitions in theoretical and empirical discovery (Q3.1-Q3.3). The background questions related to research field (B1), research orientation / interest (theoretical or empirical, B2), gender, academic position, age, language skill, and residence (B3-B7). The question types that were used were mostly 5 point Likert scale, some multiple choice questions, one ranking question, open answer questions. The answer options / statements of each question were randomized. The full survey will be made available online at https://osf.io.

3.2 Participants

Subjects were recruited from three larger academic fields: natural science, social science, and history and philosophy of science. The survey was advertised via email lists (such as Philos-L, HOPOS, PSA), social media (Twitter and Facebook), and by contacting science department heads and members via email. The online platform used to implement the survey was *Qualtrics*. The return was as follows: 86 HPSers, 214 natural scientists, and 278 social scientists.[6] The average age of these three groups were, 38.6, 47.8, and 49.2, respectively (see also Appendix 1). The percentage male in these groups was 80%, 32%, and 25%, respectively. See Appendix 1 for a detailed overview of the disciplines of the participating scientists. Subjects spent 782 seconds (ca. 13 min) on average on the survey.

4 Results

This section provides a summary of the results and a discussion regarding the following questions of the survey: Ranking of virtues (Q1.2), Epistemicity of virtues (Q2.2 and Q2.3), Nature of virtues (Q2.1 and Q2.4), and Aesthetics (Q2.5). Hypotheses H1 and H2 are discussed in the first section (4.1), H3 in the second section (4.2), H4 in the third section (4.5), and H5 in the fourth section (4.6).[7] H6 will be assessed in each section.

4.1 Theoretical Virtues: preferences

In order to find out about subjects' preferences regarding theoretical virtues (hypothesis H1), subjects were asked to order the five (plus one) Kuhnian virtues according to their perceived importance.

[6] A further 226 subjects begun the survey, but didn't complete the survey. Those subjects were excluded from the analysis. Since the survey was distributed through department heads who passed on the survey request to their colleagues, it is hard to estimate an accurate participation percentage. However, the return for people contacted through the email list kindly provided to us by Prof. *blinded for review* was about 5%.
[7] The results of the remaining questions will be published elsewhere.

Please rank the following statements in the order of importance to you (1 = most important, 2 = 2nd most important, etc.). You can change the position by dragging and dropping the statements.
"A good scientific theory should ... "

- *… be accurate with regard to all the relevant data.*
- *… be simple.*
- *… be externally consistent: it shouldn't contradict what other established theories say.*
- *… be internally consistent: it shouldn't contain contradictions.*
- *… accurately predict phenomena that weren't known before the theory was devised.*
- *… unify phenomena that prior to the theory appeared to be unconnected.*[8]

On the basis of the subjects' responses, the Condorcet winner was determined. The findings are summarized in Table 1.

virtue	HPS	Nat	Soc
accurate	2nd(4)	2nd (4)	2nd (4)
simple	6th (0)	5th (1)	5th (1)
externally consistent	5th (1)	6th (0)	6th (0)
internally consistent	1st (5)	1st (5)	1st (5)
prediction	3rd (3)	3rd (3)	3rd (2)
unification	4th (2)	4th (2)	3rd (2)

Table 1 Results of a Condorcet winner analysis regarding the ranking of theoretical virtues. Internal consistency was ranked first, accuracy second, and predictive power third, in all three groups (HPS, Nat, Soc). The number of pair-wise wins of a particular virtue against other virtues are in brackets.

It turns out not only that there is a majority preferring a particular virtue per group (confirming H1), but also that there is strong agreement across all three groups regarding the ranking of theoretical virtues (disconfirming H6): all groups ranked internal consistency higher than any other virtue, then accuracy, and then predictive success. Subjects in all three groups ranked simplicity very low (HPS: lowest, Soc and Nat: second lowest). H2 is therefore confirmed as well. The other virtue of interest for the realism debate, unifying power, was ranked 4th by HPSers and Natural scientists, and joint 3rd by Social Scientists.

Figure 1 plots the number of subjects ranking the six virtues, giving a higher 'resolution' of the result summarized in Table 1. The following observations are most striking: In the HPS group simplicity (blue) is ranked last by a big margin (49 vs. 23), in the Social Science group external consistency (green) is ranked last by an even bigger margin (156 vs 56), and in the Natural Science group simplicity (blue) loses narrowly to external consistency (green) for the last rank (85 vs. 73). Simplicity therefore has one 'win' (against external consistency) in both the social and the natural science group, but no win in the HPS group. Internal consistency is ranked first by all groups and is followed by accuracy (HPS: 50 vs. 36, Soc: 190 vs. 90, Nat: 122 vs. 92).

[8] Note that this statement was articulated in such a way that it would avoid any appearance of a "deflationary" conception of unification (see Section 2).

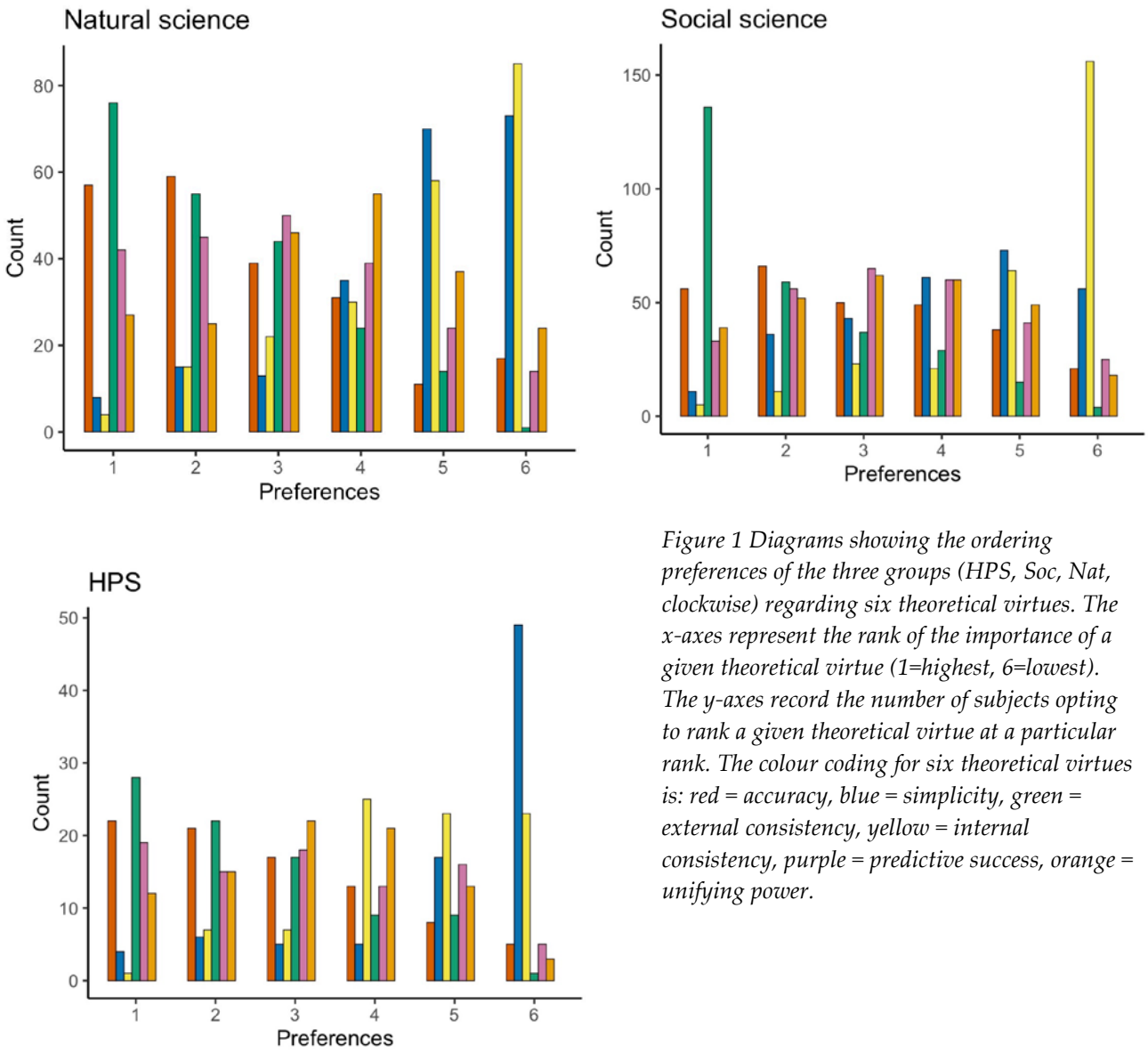

Figure 1 Diagrams showing the ordering preferences of the three groups (HPS, Soc, Nat, clockwise) regarding six theoretical virtues. The x-axes represent the rank of the importance of a given theoretical virtue (1=highest, 6=lowest). The y-axes record the number of subjects opting to rank a given theoretical virtue at a particular rank. The colour coding for six theoretical virtues is: red = accuracy, blue = simplicity, green = external consistency, yellow = internal consistency, purple = predictive success, orange = unifying power.

These results can be compared to a recent study on theoretical virtues by Mizrahi (forthcoming). Mizrahi used data mining techniques and corpus analysis to investigate the frequency of the use of theoretical virtues in the context of uses of the word 'theory', 'hypothesis', and 'model'. Mizrahi found that the most frequently mentioned virtues in all of these contexts were simplicity and consistency. Although the present survey confirms the importance of consistency, the survey's results are seemingly at odds with Mizrahi's finding that simplicity is mentioned most frequently. An explanation of that might be that scientists mention simplicity most frequently because they *because* they find simplicity problematic and therefore worthy of discussion. Mizrahi's study, however, provides little insight into why certain virtues are used more often than others (see also Section 5 for a comparison of methodologies).

As mentioned in Section 2, simplicity is widely regarded by philosophers as the most problematic theory choice criterion and it is also widely seen as non-epistemic theoretical virtue (see Section 2). Given how scientists ranked the virtues, it may well be that scientists share (either of) these concerns. The data, however, do not allow a clear verdict with regard to the question of epistemicity of simplicity (H3): it may be the case that scientists consider *also* simplicity to be an epistemic virtue, but just epistemically weaker than the other virtues.[9] Further analysis is called for (see also Section 4.2).

4.1.1 Realism and ranking

In one of the questions in the survey (Q1.1), subjects were asked about their epistemic attitudes about theories (Appendix 2). The study investigated whether subjects with a "realist" attitude would be more likely to rank simplicity and unification higher than other subjects — perhaps because they are more likely to view them as epistemic virtues. T-tests were conducted to compare the means of the rankings by realists and other subjects for each of the three groups. The results are as follows.

Realists in the social sciences ranked *unification* significantly higher than other subjects in this group ($t(278)=1.95$, $p<0.05$) and simplicity significantly lower than the other subjects in this group ($t(278)=2.59$, $p<0.01$). Realists in the natural sciences ranked *simplicity* significantly higher than the other subjects in this group ($t(212)=1.90$, $p<0.01$) and unification significantly lower than other subjects in this group ($t(212)=2.43$, $p<0.01$).[10] The differences between the mean rankings of realists and their peers were more pronounced for natural scientists than for social scientists ((i) unification: 3.56 vs. 2.8 (Nat), 3.23 vs. 3.73 (Soc), (ii) simplicity: 4.64 vs. 5.25 (Nat), 4.22 vs. 3.57 (Soc)).

In the HPS group, there were no differences between realists and their peers, both with regards to simplicity ($t(84)=0.67$) and unification ($t(84)=0.21$, $p=0.83$. The means in this group were: 3.22 vs. 3.14 (simplicity) and 4.94 vs. 5.19 (unification).

Interestingly, there wasn't any other theoretical virtue for which being realist made a significant difference to the ranking (for either social or natural scientists). Simplicity and unification have therefore rightly been identified in the realism debate as virtues worth discussing. All the stranger that there are no differences in the HPS group regarding ranking preferences between realists other subjects.

4.1.2 Professional orientation and ranking

In one background question (B2), subjects were asked about their professional orientation (theoretical or empirical, see Appendix 3). Having a theoretical orientation in one's research, one may speculate, may make it more likely to rank unification or simplicity higher than having an empirical orientation. Somebody working with mostly theoretical problems in their field, one may

[9] Schindler (2018) defends a form of realism according to which theoretical virtues such as simplicity and unification are in fact just weakly epistemic.

[10] The fact that the results are reversed for natural and social scientists explains why no differences were observed between realists and other subjects when all subjects (from all three groups) were considered.

think, may find these "internal" virtues more important than subjects working with more empirical problems. However this turned out not to be the case. T-tests of the mean rankings of simplicity and unification by theorists vs. experimentalists were all negative for all groups.

4.2 Epistemicity

In two questions of the survey (Q2.2 and Q2.3), it was further explored whether subjects perceived (philosophically controversial) theoretical virtues as epistemic or merely pragmatic (i.e., hypothesis H3).

4.2.1 Epistemicity: confidence

In the first question (Q2.2) it was determined whether subjects' confidence in a theory's correctness would be increased if a theory was simple or unified:

Please indicate your (dis-)agreement with the following statements:
"My confidence that a theory is correct is increased (everything else being equal), when ..."

- *... a theory unifies phenomena previously thought to be unrelated.*
- *... a theory is simple.*

The results of this question are depicted in Figure 2.

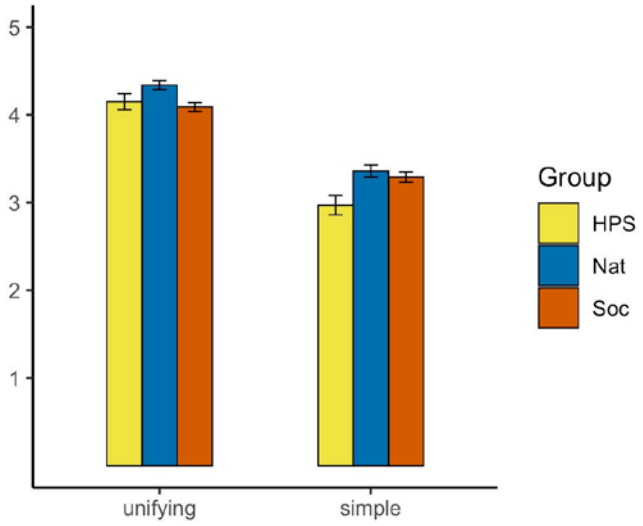

Figure 2 Subjects' views regarding whether their confidence in a theory's correctness was raised by the theory unifying the phenomena or by being simple. The y-axis represent Likert scale ratings and the x-axis the two statements (unifying, simple). Error bars are standard errors.

With regards to the first statement (unification), subjects' confidence in a theory's correctness is robustly increased for all three groups (HPS, Nat, Soc) when a theory unifies hitherto unrelated phenomena. The means on a 5-point Likert scale for the three groups are: HPS=4.15, Nat=4.34, Soc=4.09. Natural Scientists were significantly more in agreement with the first statement

(unification) than the social scientists (t(485.94)=333.53, p<0.01, Cohen's d = 0.3 (small effect size)),[11] but there wasn't any significant difference between HPSers and either science group.

As to the second statement (simplicity) there was neither strong agreement nor disagreement for any of the three groups (see Figure 2). The means for the three groups were: HPS=2.97, Nat=3.36, Soc=3.29. However, both natural scientists and social scientists were significantly more likely to agree with the second statement than HPSers (respectively: t(163.66)=288.29, p<0.01, Cohen's d = 0.36 (moderate effect size) and t(139.08)=255.37, p<0.01, Cohen's d =0.32 (small effect size).

In sum, there is evidence that subjects (from all groups) perceive unification as an epistemic virtue. There is no evidence that subjects perceive simplicity as an epistemic virtue, but there is also no evidence that they perceive simplicity as merely a pragmatic virtue (for then they should have disagreed that simplicity raises their confidence in the correctness of a theory). Hence H3 can be rejected with regard to unification, but can neither be confirmed nor disconfirmed with regard to simplicity.

4.2.2 Epistemicity: conflict

In the second question it was determined whether subjects would find it acceptable if a theory was conflicting with the data when it had other (philosophically controversial) theoretical virtues.

Please indicate your (dis-)agreement with the following statements:
"It's acceptable for a theory to be in conflict with some of the relevant data ..."

- *... when a theory unifies phenomena previously thought to be unrelated.*
- *... when a theory is simple.*
- *It's not acceptable for a theory to be in conflict with any of the relevant data.*

The rationale behind this question was that if theoretical virtues such as unification and simplicity are merely pragmatic virtues (H3), then conflicts with the data should not be acceptable, because the match with the phenomena is the overriding *epistemic* reason for accepting a theory (van Fraassen 1980, see Schindler 2018).

The results of this question are depicted in Figure 3. With regard to all three statements, all three groups gave answers around the midpoint of the 5-point Likert scale (neither agree nor disagree).[12] The means for the first statement are (unification): HPS=3.44, Nat=3.08, Soc=3.4. For the second statement (simplicity): HPS=2.73, Nat=2.54, Soc=3.21. And for the third statement (conflict not acceptable): HPS=2.64, Nat=3.1, Soc=2.55. There were some significant differences between the groups though.

[11] The effect sizes, as determined by Cohen's d, were categorized in accordance with a proposal by Hyde (2005) and Machery (2017, 46). A moderate effect size according to this classification is between 0.36 and 0.65.

[12] Throughout this paper, it was checked with t-tests (at the 99% level) whether the means of the responses of each group were smaller than 4 or bigger than 2 on the 5-point Likert scale.

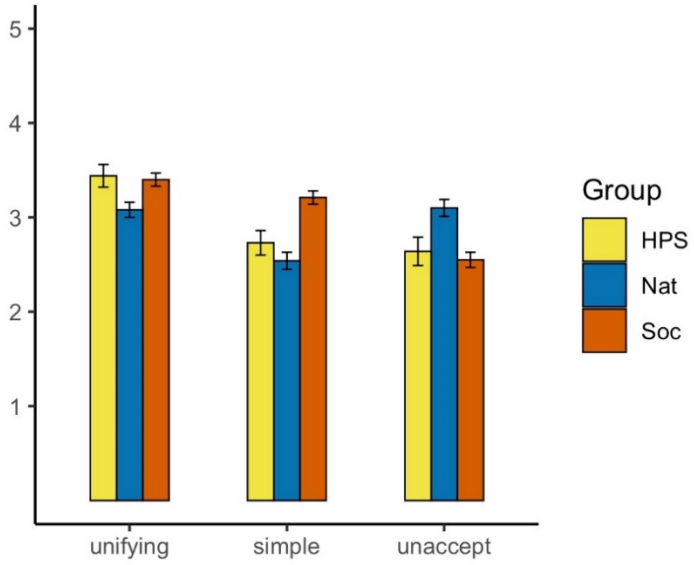

Figure 3 Subjects' views regarding the conflict between theory and data, if the theory is unifying (left) or simple (middle), and their views regarding the unacceptability of conflicts between theory and data. The y-axis represent Likert scale ratings and the x-axis the three statements (unifying, simple, and unacceptable). The error bars are standard errors.

HPSers and Social Scientists were more likely to accept conflicts between theories and the data than natural scientists, when the theory was *unifying* (respectively: t(171.72)=245.40, p=0.02, Cohen's d=0.31 (small) and t(430.98)=295.15, p<0.01, Cohen's d=0.27 (small)). Natural scientists and HPSers were less likely than social scientists to accept conflicts between the theory and data when the theory was *simple* (respectively: t(137.32)=317,401, p<0.01, Cohen's d=0.39 (large) and t(443.21)=599.43, p<0.01, Cohen's d=0.55 (moderate); social scientists responses was at the midpoint). Social scientists and HPSers were more likely than natural scientists (whose average response was at the midpoint) to *dis*agree that conflicts between theory and data are unacceptable (respectively: t(453.77)=447.70, p<0.01, Cohen's d=0.41 (moderate) and t(155.75)=264.23, p=0.01, Cohen's d=0.34 (small)). It can therefore be concluded that, from the three groups tested, social scientists seem to be most accepting of conflict between theory and data.

It is interesting to note that, despite the mean responses being around the midpoint of the 5-point Likert scale for all three statements, there was a quite sizable number of subjects in all three groups who agreed that conflicts between theory and data were acceptable when the theory is unifying (see Figure 4). No such substantive agreement was found with regard to the second or third statement. There is therefore reason to think that subjects from all three groups are more willing to treat unification than simplicity as an epistemic virtue (in accordance with the results of the test discussed in the previous section). In sum, the results for question Q2.2 indicate that H3 is to be rejected with regards to unification, but it cannot be rejected or accepted when it comes to simplicity.

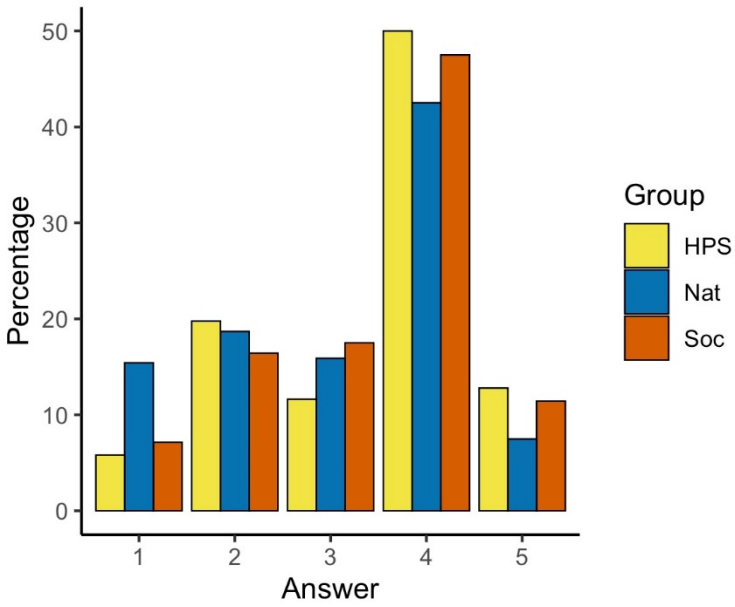

Figure 4 Substantive number of subjects agree that a unifying theory may conflict with the data: HPS=63% agree, 26% disagree, Nat=50% agree, 34% disagree, Soc =59% agree, 24% disagree.

Question Q2.2 also contained an optional open subquestion:

It's acceptable for a theory to be in conflict with some of the relevant data, when ... (optional)

A substantial number of subjects from all three groups answered this optional question: HPS (25), Nat (79), Soc (92). The responses were coded manually (see Appendix). The three most frequent response given were (in that order): D = the data may turn out to be unreliable or untrustworthy / there is reason to doubt data (30%), I = all theories are idealisations / can never explain all the data / reality is too complex (17%), and E = theory successfully explains important or most facts, and fails to accommodate only less important or fewer facts (12%). See https://osf.io for all responses (to be uploaded after publication).

4.3 Epistemicity and realism

Are realists more likely to view theoretical virtues as epistemic? To test this, a logit model was employed. The results are as follows (for details see Appendix 2).

With regards to Q2.2, subjects agreeing with the realist attitude from Q1.1 (see Appendix 2) were not more likely to have their confidence in a theory raised by either unification (HPS: p=0.208, Nat p=0.629, Soc: p=0.06) or simplicity (HPS: p=0.29, Nat: p=0.67, Soc: p=0.73).

With regards to Q2.3, HPS realists were 0.26 more likely than other HPSers to agree that conflict is acceptable when the theory unifies (p=0.03). Social science realists were 0.24 more likely than other social scientists to agree that conflict is acceptable when the theory unifies (p<0.01). There was no difference between natural science realists and other natural scientists (p=0.17).

There were no significant differences between realists and other subjects with regard to both statement 2 (simplicity) and statement 3 (conflicts are unacceptable).

4.4 Epistemicity concluded

The data show that *unification* raises subjects' confidence in a theory's correctness (in all three groups). At least a large subset of subjects in each group also finds it acceptable that a theory conflicts with the data, so long it is unifying. Overall then, there is evidence that unifying power is viewed as an epistemic virtue amongst scientists and HPSers. H3 can be rejected with regards to unifying power and H6 disconfirmed.

Both with regards to the question of whether *simplicity* raises subjects' confidence in a theory's correctness and with regards to the question of whether theories may conflict with data when theories are simple, subjects' responses were at the midpoint of the 5 point Likert scale (regardless of group). In other words, there is no evidence that subjects view simplicity as an epistemic virtue. On the other hand, given that subjects also didn't disagree that simplicity raises confidence and given that subjects also didn't disagree that it is acceptable for simple theories to be in conflict with the data, there is no evidence that subjects believe that simplicity is merely a pragmatic virtue. In other words, the survey is inconclusive with regard to the epistemic status of simplicity. H3 with regards to simplicity can neither be rejected nor accepted. Since this is true for all three groups, H6 is disconfirmed by this result

Realists in the three groups are not more confident in the correctness of a theory if the theory is either unifying or simple than other subjects. Realists in the social sciences and in HPS (but not in the natural sciences) were more likely to accept conflict with the data when the theory is unifying (but not when it was simple). When seen in light of the other results, one may say that "non-realists" were as much convinced of the confidence-raising abilities of unification as the realists, and as indifferent about the confidence-raising abilities of simplicity as the realists. Realists in the social science and in HPS, however, *were* more accepting of conflicts with the data when the theory was unifying than other subjects (but, again, they were just as indifferent about simplicity).

4.5 Nature of virtues

There were two questions in the survey asking subjects about the nature of virtues (relating to H4). In one of these questions the focus was on the nature of simplicity, and the other asked about the relationship between simplicity and unification.

4.5.1 Simplicity

Simplicity is a notoriously vague virtue; it can come in many shapes and forms. As mentioned in Section 2, there are two broad forms have however been identified, namely syntactic and ontological parsimony (Baker 2016). When a theory is syntactically parsimonious, it employs relatively few theoretical principles in accommodating the phenomena. When a theory is ontologically parsimonious, it employs a relatively small number of entities in accommodating the phenomena. There are many further questions and complications, such as 'how few principles or entities must a theory employ in order to be simple?' and 'what is the relevant (sub)form of simplicity for theory choice?', 'why should simplicity (in either form) be truth-conducive?', etc. (Baker 2016, Achinstein 2018, Schindler 2018). However the broad distinction between ontological and syntactic simplicity is clear enough.

There is yet a further, more technical sense of simplicity that has been discussed in the philosophy of science literature. It concerns simplicity in terms of parameter freedom (Forster and Sober 1994, Sober 2015). A free parameter in a theory is a parameter whose value is not determined theoretically, but need to be 'fixed' on the basis of experiments. An well-known example is the standard model in particle physics, which doesn't predict many of the masses of the particles that figure in it (Friederich et al. 2014). Limited parameter freedom is widely regarded as a virtue, since it is harder to accommodate the phenomena in an ad hoc fashion (Worrall 1989, Forster and Sober 1994, Hitchcock and Sober 2004, Worrall 2014, Sober 2015, Schindler 2018). And the fewer the free parameters, the 'simpler' the theory or model.

Given these three basic forms of simplicity, the survey posed the following question:

What makes a theory simple for you? A theory is simple, if ...

- *... it has a small number of free parameters.*
- *... it uses a small number of basic principles.*
- *... the number of basic things which it postulates (e.g. particles) is small.*

The results of this question are depicted in Figure 5.

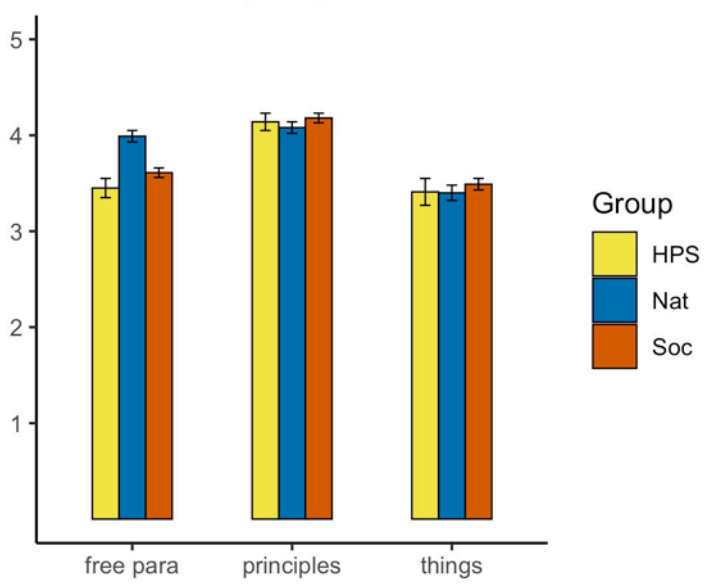

Figure 5 Subjects' views regarding the nature of simplicity. The y-axis represent Likert scale ratings and the x-axis the three statements (few free parameters, few principles, few things). Error bars are standard errors.

Natural scientists were significantly more likely to agree with the first statement (simplicity as paucity of parameter freedom) than HPSers and social scientists: $t(154.31)=451.47$, $p<0.01$, Cohen's $d=0.58$ (moderate) and $t(450.33)=455.75$, $p<0.01$, Cohen's $d=0.41$ (moderate), respectively. The means were: 3.45 (HPS), 3.99 (Nat), and 3.61 (Soc). That natural scientists were more sympathetic to this form of simplicity may have to do with the fact that the natural sciences are more formalized than the social sciences. Accordingly, parameter freedom is a notion that may simply be more relevant to natural scientists.

With regard to the second statement (simplicity as few basic principles), there was broad agreement amongst all three groups. The means were: 4.14 (HPS), 4.08 (Nat), and 4.18 (Soc). There were no significant group differences.

Finally, all three groups neither agreed nor disagreed that simplicity was to be understood in terms of a small number of postulated entities. The means were: 3.41 (HPS), 3.4 (Nat), and 3.49 (Soc). There were no between-group differences. This is further disconfirming evidence for H6. This 'null result' is surprising; it may indicate that ontological parsimony is not as important a consideration for scientists when they assess the simplicity of a theory than syntactic parsimony. Given that much of the philosophical literature has focused not so much on syntactic parsimony, but rather on ontological parsimony (Nolan 1997, Baker 2003, Jansson and Tallant 2017), this is an important result.

In sum, H4 (scientists prefer one form of simplicity) can be accepted with regards to syntactic parsimony and paucity of parameters (for natural scientists only). H4 is neither confirmed nor disconfirmed with regards to ontological parsimony. The results regarding ontological and syntactic parsimony disconfirm H6, but given the disagreement particularly between natural scientists and HPSers about simplicity and parameter paucity, H6 is confirmed with regard to that form of simplicity.

4.5.2 Simplicity and unification

Accounts of unification often appeal to simplicity. For example, Kitcher's influential (and still dominating) account has it that a theory is unifying if it allows one to derive a large number of phenomena from a small number of argument patterns (Kitcher 1981). In other words, the fewer argument patterns a theory employs in accommodating the phenomena, the more unifying it is. In order to check whether there is this kind of reciprocal relationship between unification and simplicity, subjects received the following prompt:

Do you agree with the following statement? The more unifying a theory, the simpler it is.

The mean results of all groups were at the midpoint of the 5-point-Likert scale (Nat=2.95, SD=1.04; Soc=2.79, SD=0.92; HPS=2.78, SD=1.12). There were no significant differences between the groups. H6 is thereby disconfirmed. The results allow for no conclusion regarding the relationship of simplicity and unification (either way).

4.6 Aesthetics

In science, theoretical virtues are often associated with aesthetic judgments (McAllister 1999). For example, Einstein's "highest praise" for a good theory reportedly was that it was "beautiful" and he even is said to have considered beauty more important than empirical accuracy (McAllister 1999, 96). In order to check how widespread aesthetic sentiments among scientists, the survey gave subjects the following prompt:

Please indicate your (dis-)agreement with the following statement:
"A theory can be beautiful', 'elegant', or 'appealing' when it ..."

- *... is simple.*

- *... is unifying.*
- *... is not ad hoc.*
- *... has not many free parameters.*
- *I generally don't consider theories aesthetically appealing.*

The results for this question are depicted in Figure 6. The survey revealed that all groups agreed with the first two statements (aesthetic when simple or unifying). Only the social scientists agreed with the third statement (aesthetical when non-ad hoc) and none of the groups agreed with the fourth statement (aesthetical when few free parameters). All groups disagreed with the fifth statement (theories are not aesthetically appealing).[13] The means are summarized in Table 9 in Appendix 5. H5 is confirmed with regards to simplicity and unifying power, but not with regards to parameter freedom, and only for one group (social scientists) with regards to ad hocness. Given the wide agreement between the groups, H6 can be seen as mostly disconfirmed.

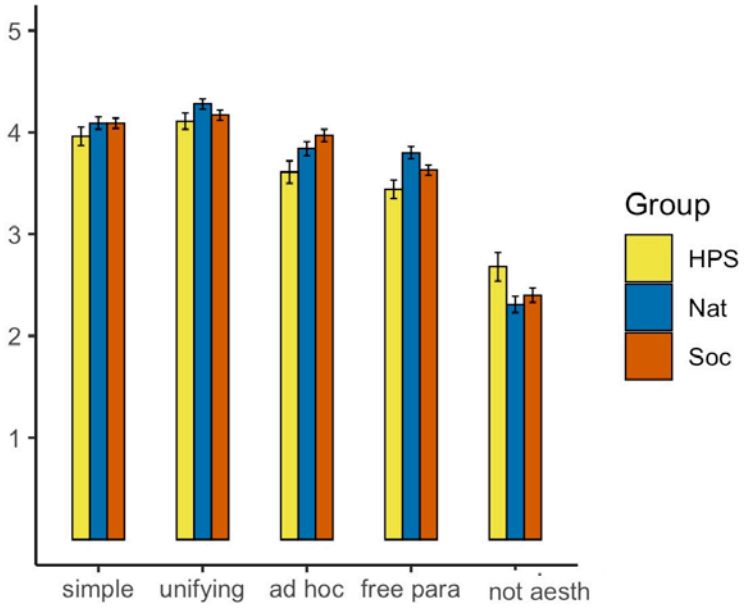

Figure 6 Subjects' views regarding the aesthetics of theoretical virtues. The y-axis represent Likert scale ratings and the x-axis the five statements (simple, unifying, not ad hoc, not aesthetically appealing). Error bars are standard errors.

There were some significant differences between the groups. Natural scientists were more likely to agree with statement 4 (theories with fewer free parameters aesthetically pleasing) than HPSers and social scientists (respectively: $t(159.71)= 328.54$, $p<0.01$, Cohen's $d = 0.42$ (moderate) and $t(453.05)=216.94$, $p<0.05$, Cohen's $d= 0.2$ (small)). This result accords with natural scientists being more likely to recognize paucity of parameter freedom as a form of simplicity (see Section 4.5.1). Social scientists were also significantly more likely than HPSers to agree with statement 3 (non-ad hocness is aesthetically pleasing): $t(134.58)=299.43$, $p<0.01$, Cohen's $d=0.38$ (moderate). Finally,

[13] Again, all this was determined by testing (with t-tests at the 99% level) whether the groups' average responses were significantly different from 4 or 2 on the 5 point Likert scale.

HPSers were more likely than social natural scientists to agree that theories are not aesthetically pleasing (t(141.13)=224.87, p<0.05, Cohen's d=0.3 (small).

Question 5.2 also contained an optional subquestion:

Are there any other circumstances in which you would say that a theory is elegant? (optional)

There were far fewer responses than to the other open question of this survey (see Section 4.2.2): HPS=12, Nat=34, Soc=32. No clear pattern emerged. See https://osf.io for all responses (to be uploaded after publication).

4.7 Aesthetics and professional orientation

One may speculate that subjects with a theoretical orientation may be more likely to find theoretical virtues aesthetically pleasing than subjects whose research has a more empirical orientation Appendix 3. T-tests showed that this was indeed the case for statement 3 (not ad hoc) in the natural sciences, and statement 1 and 2 (simplicity, unifying power) and 4 (not aesthetic) in the social sciences. The result for the natural sciences are: t(176.51)=2.12; p=0.04. The means are: 4.01 (theo) vs 3.74 (emp), Cohen's d=0.29 (small). The results for the social sciences are as follows: t(240.73)=2.22; p=0.03 (simplicity), means= 4.24 (theo) vs. 4.01 (emp), Cohens' d=0.27 (small); t(233.13)=3.09; p<0.01 (unification), means = 4.36 (theo) vs. 4.06, Cohens' d=0.38 (moderate); and t(207.97)=2.79, p=0.01 (not aesthetic), means = 2.12 (theo) vs. 2.55 (emp), Cohens' d=0.35 (small). There were no differences between theoretically oriented and empirically oriented researchers in the HPS group.

5 Weaknesses of the study and comparison

One problem with quantitative surveys such as this one is that subjects may understand the terms used in the questions differently than intended by the experimenter. Although one cannot fully rule out such misunderstandings, one can guard against them by formulating clearly, concisely, and with a minimal amount of jargon. Great effort was invested in optimizing these goals in this survey and previous versions were tested and modified after several pilots.

As mentioned briefly previously (Section 4.1), other empirical methods that have recently been applied to the topic of theoretical virtues are text mining and corpus analysis (Mizrahi forthcoming). An advantage of these methods is that many of the aforementioned difficulties of designing a survey can be avoided. But also these other methods have disadvantages. For example, as Mizrahi (forthcoming) points out himself "it is possible that scientists value theoretical virtues, which guide them in theory choice, but that scientists simply do not invoke theoretical virtues explicitly in their published work all that frequently". Indeed, Mizrahi found that only 15-30% of publications in the life, physical, and social sciences (which contains the term 'theory', 'hypothesis', or 'model') mentions *any* of the Kuhnian standard theoretical virtue. Even a virtue like accuracy, which is widely considered the most important virtue by philosophers (van Fraassen

1980, Psillos 1999) is "rarely" mentioned explicitly.[14] It is therefore probably safe to say that the proxy of explicit mentions of virtues in published papers severely underestimates the actual use of virtues in practice.

Finally, philosophers have traditionally used historical case studies to support their views concerning theoretical virtues (Schindler 2018). Many objections have been raised against the case study approach (Giere 1973, Pitt 2001, Kinzel 2015, Mizrahi 2020), but it remains a method of choice in many quarters in the philosophy of science.

With any method we choose to study science, we will have to make compromises. So long as we as a community diversity our methods, it is to be hoped that, on aggregate, we will end up with the best picture of science we can get.

6 Conclusion

This paper presented results of an exploratory survey into natural and social scientists' views regarding theorizing and theoretical virtues. It compared these views to views held by people working in the field of "history and philosophy of science" (mostly philosophers).

The survey found that scientists (and HPSers) do prefer some virtues to others (confirming H1): internal consistency was ranked first, accuracy second, and predictive success third by all groups. Simplicity was consistently ranked low, confirming H2: scientists ranked it only fifth and HPSer sixth most important (out of six virtues. Social scientists with realist leanings ranked unification higher than their colleagues and natural scientists with realist leanings ranked simplicity higher than their peers. This may indicate that realists in these two groups attribute more epistemic weight to these two virtues than their colleagues. No differences were founds for realist HPSers and their colleagues.

Subjects from all three groups stated that their confidence in a theory's correctness is raised when the theory is unifying (everything else being equal). That wasn't the case for simplicity. There is some evidence that at least a sizable number of subjects from all groups find it acceptable that a theory conflicts with the data, so long as the theory unifies the phenomena. There was no such evidence for simple theories and conflicts with the data. This disconfirms H3 with regards to unification, but neither confirms nor disconfirms H3 with regards to simplicity. Since there is agreement across the groups (both with regard to the (lack of) epistemicity of simplicity and unification), H6 is disconfirmed. Realist HPSers and realist social scientists (but not natural scientists) were more likely than their colleagues to agree that conflicts with unified theories are

[14] Mizrahi discusses only the most frequently mentioned virtues per discipline. In the context of "model talk", accuracy is the most frequently occurring virtue in the life sciences and (together with consistency) in the social sciences (5-17%). Accuracy does not figure among the most frequently mentioned virtues in the context of theory and hypothesis talk. Another disadvantage of the methods used by Mizrahi were mentioned in Section 4.1: they are very limited for detecting the *reasons for why* certain virtues are mentioned (or not) in published papers.

acceptable. Realists did not differ from their colleagues with regard to confidence raising of theories.

As to the broader implications for the scientific realism debate, philosophers may draw the following conclusions from this survey. First of all, scientists do seem to view unifying power as genuinely *epistemic* virtue. This contrasts with the most prominent antirealist view defended by van Fraassen (1980), but is in agreement with the views exhibited by the HPSers in this survey. With regards to the simplicity, the results of this survey are inconclusive: scientists did not agree nor disagree that simplicity raises their confidence. Had simplicity been a pragmatic virtue for scientists, they should have disagreed that is does.

The survey found agreement in all three groups that theories with few basic principles are simple. Natural scientists also agreed that theories with few free parameters are simple. Somewhat surprisingly, none of the three groups agreed (or disagreed) that a theory is simple if it postulates a limited number of basic entities. Philosophical discussions therefore may have to focus more on syntactic parsimony rather than (as hitherto) on ontological parsimony. These results confirm H4 with regard to syntactic parsimony and parameter freedom (for natural scientists only), but not with regard to ontological parsimony. Since there was agreement across the groups for all three forms of simplicity, H6 received further disconfirmation. The results of the survey allow no conclusions regarding the relationship between unification and simplicity.

All three groups agreed that simple and unifying theories are aesthetically appealing and disagreed that parameter freedom was aesthetically appealing. Only one group agreed that non-ad hocness was aesthetically appealing, namely social scientists. All groups disagreed that theories were not at all aesthetically appealing. H5 is therefore mostly confirmed. Given the wide agreement across the groups, H6 can be considered disconfirmed. Researchers from the natural sciences with a theoretical orientation found non-ad hocness more aesthetically appealing than researchers with an empirical orientation. Researchers from the social sciences with a theoretical orientation found simplicity and unification more appealing and rejected more clearly the view that virtues are not aethetically appealing than their more empirically oriented colleagues.

Overall, the survey revealed no major disagreements between scientists and HPSers (disconfirming H6), which is of course good news.

Appendix 1

Disciplines of participating subjects (in descending order):

- **Social Science**: economics (140), linguistics (41), political science (39), sociology (31), psychology (20), info not provided (1)
- **Natural Science:** physics (103), biology (70), chemistry (23), geology (17)

Career stage

	Tenured	Tenure-track / postdoc	PhD student	Other
HPSers	17%	26%	34%	15%
Natural Scientists	48%	20%	18%	14%
Social Scientists	52%	20%	13%	14%

The "Other" category comprises mostly retired or non-tenured academics.

Language

	Native English speakers	Advanced English speakers
HPSers	43%	55%
Natural Scientists	46%	48%
Social Scientists	55%	41%
Overall	50%	46%

Country of residence (in descending order):

USA: 270, Denmark: 54, Sweden 43, Germany 21, Finland 18, Norway 17. Subjects from other countries (<10), in alphabetical order: Albania, Australia, Austria, Azerbaijan, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, France, Greece, India, Indonesia, Iran, Ireland, Israel, Italy, Japan, Lithuania, Luxembourg, Mexico, Netherlands, Poland, Portugal, Russia, Serbia, South Africa, South Korea, Spain, Switzerland, Turkey.

Appendix 2

The first question of the survey (Q1.1) asked subjects about their epistemic attitudes toward the best theories in their academic field:

What do you think is the case of the best theories in your scientific field?

- *They do a good job of summarizing and organizing the facts.*
- *They help us to get 'behind' the phenomena to the underlying causes or laws.*
- *They set constraints that the phenomena must obey.*

The first statement was associated with an instrumentalist view and the second statement with a realist view. Obviously, statement 1 and 2 are not exclusive: one can hold both that a theory organizes the facts and that it latches onto something real. The third statement was motivated by

more recent literature on "explanation by constraint" according to which theories explain by putting necessary constraints on the laws of nature (Lange 2017).

There was agreement with the first and the second statement among all three groups at about an equal level, indicating that the subjects held mostly realist attitudes. The means were HPS = 3.92; Nat = 4.23; Soc = 4.05 and HPS = 3.91, Nat = 4.46, Soc = 4.35, respectively. This finding accords with what was found by Beebe and Dellsén (2020).[15] The number of subjects agreeing with the realist statement 2 are recorded in

	Nat	Soc	HPS	All
Realists	194	243	65	502
Other	20	37	21	78

Table 2. This information was used in the predictive statistics discussed in Section 4.1.1 and 4.3.

With regard to statement 3, i.e., the setting of necessary constraints by theories, there was neither agreement nor disagreement within the three groups. The results concerning statement 3 were not used in this paper.

	Nat	Soc	HPS	All
Realists	194	243	65	502
Other	20	37	21	78

Table 2 Number of realists and other subjects (as determined by Q1.1).

Appendix 3

The second background question of the survey (B2) asked scientists about the orientation of their research (theoretical or empirical):

How would you best describe your own work?

- *Most of my work is of a theoretical nature.*
- *Most of my work is of an empirical nature.*

Subjects in the HPS group received a modified question, asking about subjects' interests:

What part of science that you study is of most interest to you?

- *The theoretical part of science.*
- *The empirical part of science.*

HPSers were also given the following instruction:

- *In what follows, please consider the scientific field that you know best (physics, biology, chemistry, etc.), and answer accordingly.*

The results are summarized in Table 3.

Work of …	Nat	Soc	HPS

[15] A study by Robinson et al. (2019) surveyed scientists about their views regarding what is often referred to as 'metaphysical realism' (amongst other things). See also Beebe and Dellsén (2020) for a discussion.

Theoretical nature	34%	35%	87%
Empirical nature	66%	65%	13%

Table 3 Percentages of subjects' stated orientation / interest of their research.

Appendix 4

Logit models for whether realism predicts confidence in theories (see Section 4.3)

	HPS	Natural Sciences	Social Sciences
Realism	0.80 (0.64)	0.32 (0.67)	0.73 (0.40)
Constant	1.16* (0.51)	1.73* (0.63)	0.86 (0.36)
Log Likelihood	-36	-77	-133
Random Effects Parameter (Variance)			
Field	N/A	N/A	N/A

*Table 4. DV: confidence raising (unification). Model 1 is a multilevel regression model with field of study level. Note: * represents p<0.05. Standard error in brackets. Realism measures the correct responses to the first statement of Q1.1.*

	HPS	Natural Sciences	Social Sciences
Realism	0.91 (0.61)	0.93 (0.51)	0.04 (0.35)
Constant	1.45* (0.55)	0.85 (0.51)	0.05 (0.33)
Log Likelihood	-53	-147	-194
Random Effects Parameter (Variance)			
Field	N/A	N/A	N/A

*Table 5. DV: confidence raising (simplicity). Model 1 is a multilevel regression model with field of study level. Note: * represents p<0.05. Standard error in brackets. Realism measures the correct responses to the first statement of Q1.1.*

Logit models for whether realism predicts acceptance of conflicts with data (see Section 4.3)

	HPS	Natural Sciences	Social Sciences
Realism	1.10* (0.52)	0.68 (0.49)	1.00** (0.36)
Constant	0.29 (0.44)	0.62 (0.47)	0.45 (0.34)
Log Likelihood	-54	-147	-186
Random Effects Parameter (Variance)			
Field	N/A	N/A	N/A

*Table 6. DV: conflict (unification). Model 1 is a multilevel regression model with field of study level. Note: * represents p<0.05, ** p<0.01. Standard error in brackets. Realism measures the correct responses to the first statement of Q1.1.*

	HPS	Natural Sciences	Social Sciences
Realism	0.05 (0.53)	0.46 (0.58)	0.38 (0.36)
Constant	0.69 (0.46)	1.39* (0.56)	0.27 (0.33)
Log Likelihood	-54	-126	-194
Random Effects Parameter (Variance)			
Field	N/A	N/A	N/A

*Table 7. DV: conflict (simplicity). Model 1 is a multilevel regression model with field of study level. Note: * represents p<0.05, ** p<0.01. Standard error in brackets. Realism measures the correct responses to the first statement of Q1.1.*

	HPS	Natural Sciences	Social Sciences
Realism	0.33 (0.59)	0.12 (0.72)	0.45 (0.47)
Constant	0.74* (0.27)	0.12 (0.14)	1.01* (0.14)
Log Likelihood	-51	-140	-156
Random Effects Parameter (Variance)			
Field	N/A	N/A	N/A

*Table 8. DV: conflict (none acceptable). Model 1 is a multilevel regression model with field of study level. Note: * represents p<0.05, ** p<0.01. Standard error in brackets. Realism measures the correct responses to the first statement of Q1.1.*

Appendix 5

	HPS	Nat	Soc
Simple	3.96	4.09	4.09
Unifying	4.11	4.28	4.17
Not ad hoc	3.61	3.84	3.97
Few free parameters	3.44	3.8	3.63
Not aesthetic	2.68	2.31	2.4

Table 9 Mean responses to Question 2.5 regarding the aesthetics of theories.

Appendix 6

Coding of open responses to Q2.3 (see Section 4.2.2)

- C = there are (potential) confounders which, when taken into account, can explain the conflict with the data away
- D = the data may turn out to be unreliable / untrustworthy / there is reason to doubt data
- E = theory successfully accommodates important facts / most facts, and fails to accommodate only less important / fewer facts
- I = all theories are idealisations / can never explain all the data / reality is too complex
- N = the theory is relatively new and still under development / and will have to be improved (in order to later accommodate the data)

- P = the theory is probabilistic (and doesn't need to accommodate all data)
- R = data are actually not relevant for the theory
- S = accommodating the additional data would "complicate" the theory to an unacceptable extend (this could mean that a theory could accommodate the data, but only in an unacceptable ad hoc fashion, so that the data conflict is the lesser evil)
- T = the theory is still better than the alternative theories – despite the conflict
- ? = responses were hard to make sense of

	HPS	Nat	Soc	Sum	% of all responses
C	4	4	8	16	6%
D	1+4*	38+3*	20	66	31%
E	4	11	10	25	12%
I	6	11	18	35	17%
N	3	11	5	19	9%
P	0	0	2	2	1%
R	3	3	4	10	5%
S	0	1	2	3	1%
T	0	4	14	18	9%
?	3	8	10	21	10%
sum	28	90	93	215	100%

*Table 10 Subjects stating reasons for when they consider conflict between theory and data is acceptable. Some subjects mentioned more than one reason (the number of subjects providing responses relative to the overall groups sizes were: HPS=25/86, Nat=79/214, Soc=92/278). * = the plus-sign relates to entries that mentioned the faulty interpretation of data or the flawed method of generating the data.*

References

Achinstein, Peter. 2018. *Speculation: Within and about Science*: Oxford University Press.

Baker, Alan. 2003. Quantitative parsimony and explanatory power. *The British journal for the philosophy of science*, **54** (2): 245-259.

———. 2016. Simplicity. *The Stanford Encyclopedia of Philosophy (Winter 2016 Edition)*, edited by Edward N. Zalta, <http://plato.stanford.edu/archives/fall2013/entries/simplicity/>.

Barnes, Eric C. 1995. Inference to the loveliest explanation. *Synthese*, **103** (2): 251-277.

Beebe, James and Finnur Dellsén. 2020. Scientific Realism in the Wild: An Empirical Study of Seven Sciences and HPS *Philosophy of Science*, **87** (2): 336-364.
Chall, Cristin, Martin King, Peter Mättig, and Michael Stöltzner. 2019. From a boson to the standard model Higgs: a case study in confirmation and model dynamics. *Synthese*: 1-33.
Douglas, Heather. 2009. *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press.
———. 2014. The Value of Cognitive Values. *Philosophy of Science*, **80** (5): 796-806.
Douven, Igor and Jonah N. Schupbach. 2015. The role of explanatory considerations in updating. *Cognition*, **142**: 299-311.
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*, **191** (16): 3897-3917.
Giere, Ronald N. 1973. History and philosophy of science: Marriage of convenience or intimate relationship. *British Journal for the Philosophy of Science*, **24** (3): 282-297.
Hacking, Ian. 1982. Experimentation and scientific realism. *Philosophical Topics*, **13** (1): 71-87.
Harman, Gilbert H. 1965. The inference to the best explanation. *The philosophical review*, **74** (1): 88-95.
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.
Hyde, Janet Shibley. 2005. The gender similarities hypothesis. *American psychologist*, **60** (6): 581.
Ivani, Silvia. 2018. What we (should) talk about when we talk about fruitfulness. *European Journal for Philosophy of Science*, **9** (1): 4. https://doi.org/10.1007/s13194-018-0231-7.
Ivanova, M. and S. French. 2020. *The Aesthetics of Science: Beauty, Imagination and Understanding*: Taylor & Francis.
Ivanova, Milena. 2017. Aesthetic values in science. *Philosophy Compass*, **12** (10): e12433. https://onlinelibrary.wiley.com/doi/abs/10.1111/phc3.12433.
Jansson, Lina and Jonathan Tallant. 2017. Quantitative Parsimony: Probably for the Better. *The British Journal for the Philosophy of Science*, **68** (1): 781–803. http://bjps.oxfordjournals.org/content/early/2016/02/22/bjps.axv064.abstract.
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.
Kitcher, Philip. 1981. Explanatory unification. *Philosophy of Science*, **48** (4): 507-531.
Knobe, Joshua and Shaun Nichols. 2017. Experimental Philosophy. *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*, edited by Edward N. Zalta, https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/.
Kuhn, Thomas S. 1977. Objectivity, Value Judgment, and Theory Choice. In *The Essential Tension*, Chicago: University of Chicago Press, 320-333.
Lange, Marc. 2017. *Because without cause: Non-causal explanations in science and mathematics*: Oxford University Press.
Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.
Lipton, Peter. 2004. *Inference to the best explanation*. 2nd edition ed. London: Routledge.

Machery, Edouard. 2016. Experimental philosophy of science. *A companion to experimental philosophy*: 475-490.
Mättig, Peter and Michael Stöltzner. 2019. Model choice and crucial tests. On the empirical epistemology of the Higgs discovery. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **65**: 73-96.
McAllister, James W. 1999. *Beauty and revolution in science*: Cornell University Press.
McMullin, Ernan. 1976. The fertility of theory and the unit for appraisal in science. In *Essays in the Memory of Imre Lakatos*, Robert S. Cohen (ed.), Dordrecht: D. Reidel Publishing Company, 395-432.
———. 1983. Values in science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Peter Asquith and Thomas Nickles, Two: Symposia and Invited Papers 3-28.
———. 1995. Epistemic virtue and theory appraisal. In *Realism in the Sciences: proceedings of the Ernan McMullin Symposium, Leuven, 1995*, I. Douven and L. Horsten (ed.), Leuven: Leuven University Press.
Mizrahi, Moti. 2020. The Case Study Method in Philosophy of Science: An Empirical Study. *Perspectives on Science*, **28** (1): 63-88.
———. forthcoming. Theoretical Virtues in Scientific Practice: An Empirical Study. *British Journal for the Philosophy of Science*.
Morrison, Margaret. 2000. *Unifying scientific theories*. Cambridge: Cambridge University Press.
Nolan, Daniel. 1997. Quantitative parsimony. *The British Journal for the Philosophy of Science*, **48** (3): 329-343.
Okasha, Samir. 2011. Theory choice and social choice: Kuhn versus Arrow. *Mind*, **120** (477): 83-115.
Pitt, Joseph C. 2001. The dilemma of case studies: toward a Heraclitian philosophy of science. *Perspectives on Science*, **9** (4): 373-382.
Psillos, Stathis. 1999. *Scientific realism: How science tracks truth*. London: Routledge.
Robinson, Brian, Chad Gonnerman, and Michael O'Rourke. 2019. Experimental Philosophy of Science and Philosophical Differences across the Sciences. *Philosophy of Science*, **86** (3): 551-576.
Schindler, Samuel. 2017. Theoretical fertility McMullin-style. *European Journal for Philosophy of Science*, **7** (1): 151-173.
———. 2018. *Theoretical Virtues in Science: Uncovering Reality Through Theory*. Cambridge: Cambridge University Press.
Schupbach, Jonah N. 2017. Experimental Explication. *Philosophy and Phenomenological Research*, **94** (3): 672–710.
Schupbach, Jonah N. 2011. Comparing Probabilistic Measures of Explanatory Power. *Philosophy of Science*, **78** (5): 813-829.
Sober, Elliott. 2015. *Ockham's Razors*. Cambridge: Cambridge University Press.
Steel, Daniel. 2010. Epistemic Values and the Argument from Inductive Risk. *Philosophy of science*, **77** (1): 14-34.
Steel, Daniel, Chad Gonnerman, and Michael O'Rourke. 2017. Scientists' attitudes on science and values: Case studies and survey methods in philosophy of science. *Studies in History and Philosophy of Science Part A*, **63**: 22-30.
Stotz, Karola. 2009. Philosophy in the trenches: from naturalized to experimental philosophy (of science). *Studies in History and Philosophy of Science Part A*, **40** (2): 225-226.

Stotz, Karola and Paul Griffiths. 2004. Genes: Philosophical analyses put to the test. *History and Philosophy of the Life Sciences*: 5-28.

Stotz, Karola, Paul E Griffiths, and Rob Knight. 2004. How biologists conceptualize genes: an empirical study. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **35** (4): 647-673.

van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.

van Fraassen, Bas C. 1989. *Laws and symmetry*. Oxford: Clarendon Press.

Woodward, James. 2014. Simplicity in the best systems account of laws of nature. *The British Journal for the Philosophy of Science*, **65** (1): 91-123.

Worrall, John. 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.

———. 2014. Prediction and accommodation revisited. *Studies in History and Philosophy of Science Part A*, **45**: 54-61.

Wray, K Brad. 2018. *Resisting scientific realism*: Cambridge University Press.