Experimental Philosophy of Science and Philosophical Differences across the Sciences

Abstract

This paper contributes to the underdeveloped field of experimental philosophy of science. We examine variability in the philosophical views of scientists. Using data from Toolbox Dialogue Initiative, we analyze scientists' responses to prompts on philosophical issues (methodology, confirmation, values, reality, reductionism, and motivation for scientific research) to assess variance in the philosophical views of physical scientists, life scientists, and social and behavioral scientists. We find six prompts about which differences arose, with several more that look promising for future research. We then evaluate the difference between the natural and social sciences and the challenge of interdisciplinary integration across scientific branches.

Keywords: Experimental philosophy, philosophy of science, integration, interdisciplinarity

Brian Robinson
(*corresponding author*)
Assistant Professor
Texas A&M University-Kingsville
Department of History, Political Science & Philosophy
700 University Blvd. MSC 165
Kingsville, TX 78363
brian.robinson@tamuk.edu
361-593-3602
ORCID: 0000-0003-0694-3328

Chad Gonnerman
Assistant Professor

University of Southern Indiana
Department of Philosophy
8600 University Blvd.
Evansville, IN 47712

Michael O'Rourke
Professor
Michigan State University
Department of Philosophy
368 Farm Lane Room 503
East Lansing MI 48824-1032
ORCID: 0000-0002-4629-0811

Acknowledgements: We would like to thank Stephen Crowley and Daniel Steel for discussion of the issues addressed in this article, as well as audiences at the 2015 Buffalo Experimental Philosophy Conference, the 2015 Science of Team Science Conference, and the 2016 SRPoiSE/Values in Medicine, Science and Technology Conference.

Funding: The work reported in this paper was supported by the National Science Foundation under Grant No. SES-0823058. O'Rourke's work on this project was supported by the USDA National Institute of Food and Agriculture, Hatch project MICL02261.

1 Introduction

Experimental philosophy has made great strides in several areas, including epistemology and ethics, but one area that has received less attention is philosophy of science. Of the scant experimental philosophy of science produced so far, the focus has been primarily on specific scientific concepts, entities, or properties, such as genes (Stotz, Griffiths, and Knight 2004;

Griffiths and Stotz 2008) or innateness (Griffitsh, Machery, and Linquist 2009; Knobe and Samuels 2013). (For a review, see (Machery 2016).) Already a clarion call has sounded for more experimental work in philosophy of science (Stotz 2009b, 2009a; Weinberg and Crowley 2009; Nagatsu 2013; Machery 2016). We concur and also believe that the scope of this work should be extended. To these ends, we endeavor to develop and extend experimental philosophy of science by engaging in philosophical cartography. Using data gathered by the Toolbox Dialogue Initiative (TDI),[1] we attempt to illuminate some of the general contours of a map of the philosophical attitudes of practicing scientists and then use this map to contribute to current work in the philosophy of science on interdisciplinary integration.

An experimentally informed map of philosophical attitudes of scientists is philosophically interesting for several reasons. First, distinguishing between the natural and social sciences has been a topic of discussion in philosophy of science for decades now. Ian Hacking (1996), for example, offers a complex distinction between the two, claiming that the classificatory practices of the social sciences interact with the world in a way not found in the natural sciences. Call something an electron and it makes no differences to it; diagnosing people with narcissism can change how they behave. (For more on Hacking's proposal, see (Martínez 2009).) Developing a map of the philosophical attitudes of scientists can inform projects like Hacking's. For instance, if the classificatory practices of the social and natural

---

[1] Formerly the Toolbox Project (http://tdi.msu.edu/).

sciences work as Hacking proposes, we should expect social scientists to have more antirealist sympathies than natural scientists. This is a testable prediction. If it were to bear out, and thus make an appearance on our collective map of scientists' philosophical attitudes, Hacking's proposal would gain a consideration in its favor. In this paper, we develop the beginnings of such a map. We do so experimentally by basing our claims on a survey of scientists' own reports of their attitudes regarding various issues, rather than by appeal to our own intuitions or a philosophical analysis of the literatures of various sciences. This survey-driven approach allows us to capture heterogeneity across different areas of sciences, something that can elude traditional intuition- and literature-based methods in philosophy of science (Faust and Meehl 2002; Griffiths and Stotz 2008; Machery and Cohen 2012; Steel, Gonnerman, and O'Rourke 2017).[2]

Second, experimental examination of scientists' philosophical attitudes opens a new avenue into the existing discussion of interdisciplinary integration in philosophy of science. One prominent focus of philosophers of science has been on integration as a relation between various scientific objects, such as theories, fields, or specialties (Bechtel 1993; Darden and

---

[2] The traditional literature-based methods that we have in mind are ones where the philosopher engages very deeply with a subset of the scientific literature, especially literature which she deems relevant and read in depth. As Machery (2016) explains, bibliometric and cliometric techniques provide philosophers with other ways of exploring scientific literatures.

Maull 1977; Gerson 2013; Grantham 2004; Griesemer 2013). Nagel's (1961)reductionist account of scientific theories can be seen as an early entry into this literature. More recently, as skepticism has grown about this reductive project, some (Mitchell et al. 1997; Brigandt 2013; Love 2008) have proposed a non-reductive conception of integration according to which the ideas and explanations of various disciplines are integrated "to yield an overall explanation of a complex phenomenon" (Brigandt 2010, 297).

Drawing inputs from different disciplines provides reason to believe that integration of this sort is distinctly philosophical in character (O'Rourke, Crowley, and Gonnerman 2016). Disciplines engage with the world in different ways, foregrounding certain problems while leaving others in the background. This sort of engagement can be modeled in terms of a discipline's *research worldview*, comprised of a system of core beliefs, values, and practices (O'Rourke and Crowley 2013). These worldviews differ in various ways, such as in the meaning of terms used to articulate them (Bracken and Oughton 2006; S. M. Donovan, O'Rourke, and Looney 2015). The task of interdisciplinary integration typically requires combining inputs that reflect these worldviews, and differences among worldviews can complicate this task, especially since they can be fundamental and hard to see. As Eigenbrode et al. explain, "Scientists collaborating within disciplines tend to share fundamental assumptions and values concerning the scientific process and, habitually, may discuss them little, but the failure to understand and address these fundamental differences can impede progress in cross-disciplinary efforts" (2007, 56). Since the fundamental assumptions that compose disciplinary worldviews are partly metaphysical and epistemological in nature (O'Rourke and Crowley 2013), philosophy can contribute to studies of interdisciplinarity integration by investigating these assumptions. By experimentally examining philosophical commitments or attitudes of practicing scientists, we gain a better understanding of the nature, mechanisms, and challenges of integration in interdisciplinary science.

This study explores a large data set generated by TDI (Eigenbrode et al. 2007; S. M. Donovan, O'Rourke, and Looney 2015). Since 2005, over 250 TDI (or "Toolbox") workshops have been conducted worldwide to enhance mutual understanding, communication, and collaboration among members of interdisciplinary groups. After canvassing several issues in experimental philosophy of science in section 2, we offer detailed descriptions of these Toolbox workshops and the data they generate in section 3. In section 4 we present of our results. Then in section 5 we discuss those results, emphasizing what they mean for the distinction between the natural and the social sciences and for the challenge of interdisciplinary integration in the sciences.

2 Experimental philosophy of science

Our initial step toward developing experimental philosophy of science involves defending the validity and effectiveness of empirical methods in philosophy of science. Experimental philosophy is a recent and growing movement that employs the systematic collection and analysis of empirical data as means for philosophical investigation. A central tenet of this movement is that philosophical arguments often make, entail, or presuppose empirical claims, and, as such, the strength of these arguments can be judged, in part, by collecting and analyzing empirical data that bear on those claims. Experimental philosophers have moved into a wide range of philosophical debates in metaphysics, ethics, and epistemology (Knobe et al. 2012; Sytsma and Livengood 2015; Machery 2017). Yet almost ironically, the field of experimental philosophy of science has received little development (Machery 2016).

Stotz (2009a, 2009b) and Weinberg and Crowley (2009) outline some of the basic theoretical motivations for experimental philosophy of science. They assert that philosophy of science should be informed by how scientists conceive of and practice science. For this reason—as most contemporary philosophers would agree—philosophy of science that is conducted purely "from the armchair" with little awareness of actual scientists, scientific practices, and scientific findings is, at least prima facie, of dubious value. In our view, this common starting point for much of contemporary philosophy of science paves the way for the use of empirical data in philosophy of science.

The idea that philosophy of science can profit from empirical data about science is hardly novel. It is exemplified in the work of those who take what Bird (2008) calls "the historical turn in the philosophy of science." We also see it in works that use case studies to advance philosophical claims about science or its areas such as biology. A central distinction between experimental philosophy of science and these approaches, however, is the proposal that philosophy of science can also benefit from empirical analysis about how *scientists* think and behave. As Thagard notes, “Science is a human enterprise, and understanding the development of scientific knowledge depends on an account of the thought processes of humans” (1988, 4). A central focus of experimental philosophy of science, then, should be the systematic collection and analysis of empirical data pertaining to scientific thought processes and practices. And where the existing empirical record about how scientists think and behave is too thin, philosophers should feel free to gather the data themselves, using appropriate methods, of course, and possibly in collaboration with practicing scientists.[3]

In addition to data regarding the views and practices of actual scientists, another source of data could be relevant. Feist notes that “Scientific thought and behavior are not limited to scientists per se but also encompass thought processes of children, adolescents, and adults who are simulating scientific problem solving and developing mental models of how the world works” (Fiest 2006, 4). Experimental philosophy can gather and interpret data from either

---

[3] Work exists in nearby orbits, such as the cognitive sciences of science (Carruthers, Stich, and Siegal 2002; Proctor and Capaldi 2012; Feist and Gorman 2013). Some of this work draws philosophical conclusions about science from empirical data, systematically collected and analyzed (Grover 1981; Houts 1989; Wagenknecht, Nersessian, and Andersen 2015). For us, such work qualifies as experimental philosophy of science, even if the researchers don’t use the term.

trained scientists or ordinary folk without specialized scientific training. This difference in data source corresponds to the difference between experimental philosophy of science narrowly or broadly construed. Both kinds of research are represented in what little experimental philosophy of science currently exists. As examples of narrow experimental philosophy of science, Stotz et al. (2004) and Griffiths and Stotz (2008) examine scientists' conceptions of genes. Griffiths et al. (2009) and Linquist et al. (2011), alternatively, engage in a broader form of experimental philosophy of science by examining views on innateness and human nature (respectively) held by non-scientists. Knobe and Samuels (2013) combine the two by examining the views on innateness held by both scientists and non-scientists.

In all of these manifestations, the commitment of experimental philosophy of science to the relevance of empirical data is fundamentally naturalistic. Despite debate about its details, the naturalist idea that the methods of philosophy are of a piece with the methods of science figures importantly into the background of these views (Papineau 2009; Hartner 2013). That *experimental* philosophy of science is naturalistic should come as no surprise; what might be surprising, though, is the claim that philosophical aspects of science should be sensitive to data on the attitudes of scientists, *including their philosophical attitudes*. We contend that, without these data, philosophical claims about, for example, the roles of values in science or the epistemic standing of various methods will ignore a critical determinant of how values inflect actual scientific judgments, decisions, and behaviors and how methods are selected in examining specific research questions (Steel, Gonnerman, and O'Rourke 2017). To some extent, data about these philosophical attitudes may be gleaned from a close acquaintance with scientists and their publications. But any such acquaintance is likely to be non-systematic and non-representative, thereby providing an insufficient basis from which to draw philosophical conclusions about science (Faust and Meehl 2002; Griffiths and Stotz 2008; Machery and Cohen 2012). A better way to uncover these attitudes is to deploy experimental methods, especially those of the social sciences. Of course, experimentally examining the philosophical attitudes of practicing scientists (specifically those relevant to philosophical reasoning about science) does not entail that philosophers should defer to the philosophical perspectives of scientists. Rather, the success of experimental philosophy of science requires combining appreciation for conceptual dimensions of science with details about how science actually works.

3 Method

We seek to expand experimental philosophy of science by engaging in it, specifically by experimentally probing for philosophical differences among scientific branches by examining the philosophical commitments of scientists who belong to those branches. To do this, we looked at 43 Toolbox workshops that were conducted with interdisciplinary research teams.[4] These workshops were conducted between March 2009 and October 2013; they had 346

---

[4] Dataset available online:

https://mfr.osf.io/render?url=https://osf.io/ytpnw/?action=download%26mode=render participants (127 female), who ranged from graduate students to senior researchers with over twenty years of research experience.

At the start of these workshops, participants receive the STEM (Science-Technology-Engineering-Mathematics) Toolbox instrument, which consists of six modules focusing on scientifically relevant philosophical issues and concludes with a page asking for demographic data. Each module contains four to five related prompts, 28 total (Looney et al. 2013)[5] in response to which participants either report agreement or disagreement on a five-point Likert scale ranging from 1 (*Disagree*) to 5 (*Agree*) or select "Don't know" or "N/A."[6] These last two options allow participants alternate responses to the Likert scale, thereby minimizing artificial clustering around the midpoint of the scale for cases of ignorance or non-applicability. The demographic page includes four numbered spaces in which participants are asked to describe their disciplinary "identity." By analyzing participants' self-reported disciplinary identity in

---

[5] Full instrument available at https://goo.gl/g2gS7q

[6] Each module concludes with a "similar views" prompt, which asks participants to evaluate how similar the views among the group are for that module. We have excluded these prompts from the present analysis.

conjunction with their responses to these prompts, as we explain in section 4, we can look for philosophical differences across branches of science.[7]

Two preliminary tasks were required before searching the data for philosophical differences. First, we needed to assign participants to academic disciplines. We began by examining the discipline(s) listed on the demographics page. Since these responses were open-ended, there was considerable variability in the discipline(s) listed. Some provided a single discipline; others gave several. Some responded quite generally (e.g., "biology"), while others responded very specifically (e.g., "fluvial geomorphology"). The question was how to categorize participants into disciplines and larger academic branches (groups of disciplines, such as Engineering, Arts & Humanities, and Life Sciences) given this variability. This

---

[7] Workshops start with a preamble from the workshop facilitator (a TDI member), followed by participants completing a Toolbox instrument. They then engage in a semi-structured dialogue based on the prompts, with the goal of helping them to see their common research problem through each other's eyes (Looney et al. 2013). After the dialogue, participants complete the instrument a second time to assess how their views may have shifted during the dialogue. For this study, we only examined the pre-dialogue responses, since our goal is to explore the philosophical attitudes of scientists taken as representatives of their disciplines, which we take to be better represented by their attitudes prior to the Toolbox dialogue.

presented two challenges: selecting a disciplinary taxonomy to use in assigning participants to disciplines and interpreting participant responses in light of that taxonomy.

For the first challenge, we selected the Digital Commons Three-Tiered Taxonomy of Academic Disciplines (Bepress 2014) because the large degree of categorical resolution it offered made it easier to accommodate the wide variety of disciplinary self-identifications reported by workshop participants. The Digital Commons taxonomy is regularly updated based on user feedback and developed using entries from multiple sources, such as The University of California's list of departments and the Taxonomy of Research Doctoral Programs from the National Academies. At the time of analysis, the Digital Commons taxonomy comprised over 900 unique disciplinary categories broken into three tiers that we'll refer to as academic branches, disciplines, and specializations. To illustrate, Philosophy is a discipline in the academic branch of Arts and Humanities, with specializations including Metaphysics, Feminist Philosophy, and Philosophy of Science. (The Digital Commons taxonomy's third tier of specializations was especially useful for interpreting highly specific disciplinary identifications employed by Toolbox participants in their disciplinary self-identifications.) The top-tier of academic branches for the Digital Commons taxonomy consists of the following categories: Architecture, Arts and Humanities, Business, Education, Engineering, Law, Life Sciences, Medicine & Health Sciences, Physical Sciences & Mathematics, and Social & Behavioral Sciences. For this paper, we focused only on those researchers in the scientific branches: Life

Sciences, Physical Sciences & Mathematics, and Social & Behavioral Sciences. As explained below, our focus on these branches was in part due to the number of participants in each branch.

The second challenge concerned interpreting participants' disciplinary self-identifications in light of the Digital Commons taxonomy, a task complicated by demographic pages often containing multiple disciplinary specifications. In describing their disciplinary identity, most participants (276 of 346, 82.1%) provided at least two disciplines, such as "(1) algorithmics (2) microbial ecology (3) discrete math." For the sake of categorizing participants into academic branches and disciplines, we only looked at the first discipline reported for two reasons. First, we are interested in researcher's *primary* disciplinary identity, and we operated on the assumption that participants entered this discipline first. Second, it was not possible to look beyond the first discipline for everyone, since 17.9% participants only listed one discipline. We thus opted to err on the side of caution by abstaining from drawing on secondary disciplines listed to inform judgments about disciplinary identity.

After meeting the challenges associated with the first preliminary task, we turned to the second, which was to clarify the nature of the differences that could be revealed by the STEM Toolbox instrument. This task was complicated by the number of dimensions of potential difference that the instrument was designed to reveal. The STEM instrument includes prompts such as "Objectivity implies an absence of values by the researcher," and "Validation of evidence requires replication." It was designed in an explicitly philosophical fashion, with half of its prompts classified as *epistemological* and half as *metaphysical* (Eigenbrode et al. 2007;

O'Rourke and Crowley 2013). The epistemology prompts are divided into those that concern a scientist's motivation for participating in research, those that concern the methods used in conducting research, and those that concern identifying when conclusions have been reached. The metaphysics prompts are organized into those that concern the nature of reality, the structure of reality (and in particular, reductionism and emergence), and values.[8]

The STEM instrument's prompts also distinguish between how scientists view their own work and how they view science more broadly. In doing so, the prompts help isolate important determinants of a scientist's identity and place them in the context of a broader construal of scientific practice. Of the 28 prompts that we examine, 8 are reflexive prompts that express characteristics that some regard as fundamental to their own research (e.g., "My research typically isolates the behavior of individual components of a system"), and 20 are prompts that are not limited to one's own research. Of these 20 prompts, the context of interpretation varies. For example, two focus specifically on cross-disciplinary research, six use the adjective 'scientific' explicitly to frame interpretation, and five explicitly mention research without modifying it with 'scientific'. The balance talk about aspects of research without mentioning a context. Since this instrument is called the "Scientific Research Toolbox

---

[8] The values prompts reflect concerns with values in science, and so address axiological or value-theoretic concerns. See Eigenbrode et al. (2007) for why these were included in the Metaphysics half of the instrument.

Instrument" and is deployed in workshop settings with interdisciplinary science teams, the expectation is that the participants will interpret the 20 non-reflexive prompts as pertaining to research that is scientific and that extends beyond one's own practice.

Thematically, the eight reflexive prompts address aspects of science and scientific research practices that apply across multiple research projects, such as the methods one primarily uses or the ontological character of one's subject matter, with one prompt addressing the collective project of the workshop participants (viz., "The importance of our project stems from its applied aspects"). The 20 remaining prompts are more heterogeneous. Some offer provocative analyses of concepts that some take to be central to science, such as *objectivity* or *advocacy*. A second group orients workshop participants to certain visions of scientific explanation and scientific practice (e.g., "Value-neutral scientific research is possible"). Additional prompts express views about what must be true for a practice to count as science (e.g., "Scientific research must be hypothesis driven"), while still others address the values of particular scientists (e.g., "The principal value of research stems from the potential application of the knowledge gained"). This set of 20 prompts is unified in reflecting different ways of thinking about science, including its structure and function.

In indicating their agreement or disagreement with each prompt, participants reveal (in some cases perhaps only indirectly) how they practice science or how they think science should be practiced. All the prompts (but one) abstract away from specific research contexts and projects, allowing participants to understand them in the context of science as they practice it and view it. In responding to them, participants are asked to adopt the primary disciplinary perspective they assume in their research, and because of this, the responses can be understood as articulating commitments that frame one's research approach, either affirmatively or negatively.[9] Together, the STEM instrument's prompts can be seen as partly mapping an investigator's research worldview (or at least the more cognitive aspects of it), which may be understood as a set of "more or less tacit beliefs held by researchers about what they are studying and how to study it, as well as views about the nature of the output of their inquiry"(O'Rourke and Crowley 2013, 1938). When taken together, the maps that emerge for workshop participants display the topography of agreement and disagreement across a range

---

[9] This instruction is motivated by the fact that many participants have complex disciplinary identities that could issue in different responses to the prompts, depending on which part of their identity is emphasized. (See the discussion above of the disciplines listed on the demographic page.) The workshop facilitator delivers the instruction in the workshop preamble as it is articulated in the text in order to provide consistent guidance to participants in responding to the prompts. Despite this instruction, over the next couple sections, we assume that workshop participants interpret the broader science prompts as claims about scientific research in general; however, an anonymous referee helpfully pointed out that they could interpret them as more localized claims about their own particular kinds of scientific research, e.g., social scientific research or natural scientific research. In the Discussion section, we respond to this concern.

of philosophical issues that some view as fundamental to scientific inquiry and that illuminate aspects of a scientist's identity. Each map is systematic but skeletal, highlighting a few attitudes about the philosophical dimensions of science, sometimes fairly directly (as in the broader prompts about science) and sometimes more indirectly (as in the reflexive prompts). More technically, the prompts elicit, directly or indirectly, commitments about the metaphysical, epistemological, and axiological dimensions of science that often frame how one understands and practices science.

4 Results

Using participant self-reports of disciplinary identity, two of the authors served as raters to categorize participants into academic branches, using the Digital Commons Three-Tiered Taxonomy of Academic Disciplines (Bepress 2014)[10] The two raters first agreed to exclude 9 participants for providing responses that were clearly not research disciplines (such as "grant administration"). After independently coding an academic branch for each participant, the two raters then compared their results. Cohen's Kappa determined a relatively high degree of correspondence ($\kappa = 0.82$), with the raters differing on 54 participants (out of 346). Of these,

---

[10] The raters also coded and compared academic disciplines as well as academic branches. The raters agreed on 75% of cases, but there were too many disciplines (many of which were represented by only a few participants) to allow for statistical analysis.

they resolved their disagreement for all but 13 participants, who were then excluded from further analysis.[11] The results are summarized in Table 1.

Life Sciences (LS)	134
Physical Sciences & Mathematics (PSM)	72
Social & Behavioral Sciences (SBS)	58
Engineering	34
Medicine and Health Sciences	17
Arts and Humanities	8
Education	10
Business	2

Table 1. Summary of participants' academic branches as determined by two raters.

As Table 1 indicates, Life Sciences (LS), Physical Sciences & Mathematics (PSM), and Social & Behavioral Sciences (SBS) were well represented among Toolbox participants. Since we are interested in differences among scientific branches, we focused our attention on participants in these three fields, excluding participants from other academic fields from analysis. This left us with 264 participants (95 female).

For our analysis, we ran the Kruskal–Wallis rank sum test for each of the 28 Toolbox prompts.[12] To contextualize this test, for each prompt, each of the three scientific branches had

---

[11] In each of these cases, the raters agreed that the difference was intractable and both interpretations were plausible. For instance, one participant self-identified as "neurobiology," which the raters agreed could either be Life Sciences or Medicine and Health Sciences.

[12] Since our data are not normally distributed, we did not run a series of ANOVAs.

a median score. The Kruskal–Wallis test examines the three medians for each prompt for significant differences. The null hypothesis is that the medians are equal (i.e., $\mathrm{Mdn}_{LS} = \mathrm{Mdn}_{PSM} = \mathrm{Mdn}_{SBS}$). The test includes a calculation of the probability (the *p*-value reported below) of the observed results (or more extreme results), given the null hypothesis. If the probability is low enough, then we reject the null hypothesis, meaning that at least one median of the three is significantly different than the others. A follow-up test reveals which median(s) is different. If, however, the probability is not sufficiently low, the null hypothesis is not rejected, indicating that responses to this prompt are statistically indistinguishable across the three academic branches. The results of the Kruskal–Wallis test for each prompt are summarized in Table 2.[13]

Prompt	*n*	*df*	*H*	*p*	Prompt	*n*	*df*	*H*	*p*
Motivation 1	257	2	2.02	0.37	Reality 1	256	2	11.32	3.48×10^{-3}
Motivation 2	252	2	6.25	0.04	**Reality 2***	243	2	13.08	1.45×10^{-3}
Motivation 3	253	2	8.79	0.01	Reality 3	256	2	2.17	0.34
Motivation 4	258	2	0.75	0.69	**Reality 4***	237	2	35.41	2.05×10^{-8}
Methodology 1	263	2	10.45	5.38×10^{-3}	Values 1	256	2	1.45	0.48
Methodology 2*	258	2	17.19	1.85×10^{-4}	Values 2	260	2	5.07	0.08
Methodology 3*	255	2	26.19	2.05×10^{-6}	Values 3	252	2	10.90	4.29×10^{-3}
Methodology 4*	256	2	43.86	2.99×10^{-10}	Values 4	246	2	5.41	0.07
Methodology 5	257	2	2.76	0.25	Values 5	243	2	0.41	0.81
Confirmation 1	254	2	4.79	0.09	Reductionism 1	220	2	6.64	0.04
Confirmation 2	251	2	1.36	0.51	Reductionism 2	248	2	2.59	0.27
Confirmation 3*	260	2	18.44	9.91×10^{-5}	Reductionism 3	250	2	1.42	0.49

---

[13] For each prompt, we excluded "I don't know" or "N/A" responses. As is apparent in the *n* column of Table 2, typically few participants were excluded, ranging from only 6 to 44 (out of 264 total).

Confirmation 4	252	2	2.52	0.28	Reductionism 4	256	2	6.65	0.04
Confirmation 5	255	2	3.41	0.18	Reductionism 5	256	2	6.14	0.05

Table 2. Results from Kruskal–Wallis tests for each of the Toolbox prompts. Prompts marked with asterisks (*) are significant after applying the Holm-Bonferroni method for correcting for familywise error.

Because the STEM instrument was not designed to detect *particular* differences, we should be mindful of the familywise error rate associated with multiple testing. In effect, we are not testing a single hypothesis, but rather 28 hypotheses simultaneously. The danger in testing multiple hypotheses simultaneously is that it increases the probability of a false positive—a Type I error. When testing a single hypothesis, the threshold for a significant *p*-value is often set at 0.05, representing a 5% chance for a Type I error. If, however, the same threshold of 0.05 were applied separately for each of the 28 tests here, then the familywise error rate (i.e., the likelihood of at least one false positive) is approximately 0.74. So some form of correction is called for; we employed the Holm-Bonferroni method (Holm 1979), which is a stepwise algorithm that lowers the *p*-value required to reject a null hypothesis. Using this method, we find significant results for six of the prompts (marked with an asterisk in Table 2): Methodology 2, Methodology 3, Methodology 4, Confirmation 3, Reality 2, and Reality 4. Three additional prompts were very close to significant after correction: Reality 1, Methodology 1, and Values 3.[14] Since the Holm-Bonferroni method is a fairly conservative means of correcting for familywise error, we have included these three for follow-up analysis along with the other six. But these results in particular should be taken tentatively.

The Kruskal–Wallis tests found significant or nearly significant differences among the three branches of science for nine prompts. To determine where the differences reside, for each of these prompts we conducted follow-up analyses using the Mann–Whitney–Wilcoxon (MWW) test. This is a pairwise test, comparing the medians of two groups. In this case, for each prompt, three MWW tests are conducted, one for each comparison (LS and PSM, LS and SBS, and PSM and SBS). Results of all MWW tests for the nine prompts are summarized in Table 3 (Appendix).

4.1 Reality

For the Reality prompts, the initial Kruskal–Wallis tests found significant results for Reality 2 and 4, and Reality 1 was close. All of these results were borne out by the follow-up

---

[14] The *p*-values for each of these three prompts exceeded the threshold for significance set by the Holm-Bonferroni correction by less than 0.003. We therefore included these three prompts in our follow-up analyses, where we found significant differences across the scientific branches. These results should be taken tentatively and are included primarily to *suggest* future lines of research.

MWW tests (Figure 1). Researchers in Life Sciences and Physical Sciences & Mathematics were significantly were more likely to agree with Reality 1 ("Scientific research aims to identify facts about a world independent of the investigators") than those in Social & Behavioral Sciences (LS & SBS: $U = 4776.5$, $p = 9.52 \text{ x } 10^{-04}$, $r = 0.24$ and PSM & SBS: $U = 2432.5$, $p = 0.03$, $r = 0.20$). However, Life Sciences and Physical Sciences & Mathematics were not statistically distinguishable. For Reality 2 ("Scientific claims need not represent objective reality to be useful"), only between Life Sciences and Social & Behavioral Sciences was the difference statistically significant ($U = 2364.5$, $p = 3.79 \text{ x } 10^{-04}$, $r = 0.26$). As Figure 1 shows, the mean for Physical Sciences & Mathematics was between the other two, although its differences with the other two branches were not statistically significant. For Reality 4 ("The subject of my research is a human construction"), life scientists and physical scientists and mathematicians were less willing to agree with this prompt than were social and behavioral scientists (LS & SBS: $U = 1698.5$, $p = 2.70 \text{ x } 10^{-09}$, $r = 0.43$ and PSM & LS: $U = 875$, $p = 2.90 \text{ x } 10^{-06}$, $r = 0.42$). Again, Life Sciences and Physical Sciences & Mathematics were statistically indistinguishable, which is a pattern repeated below.

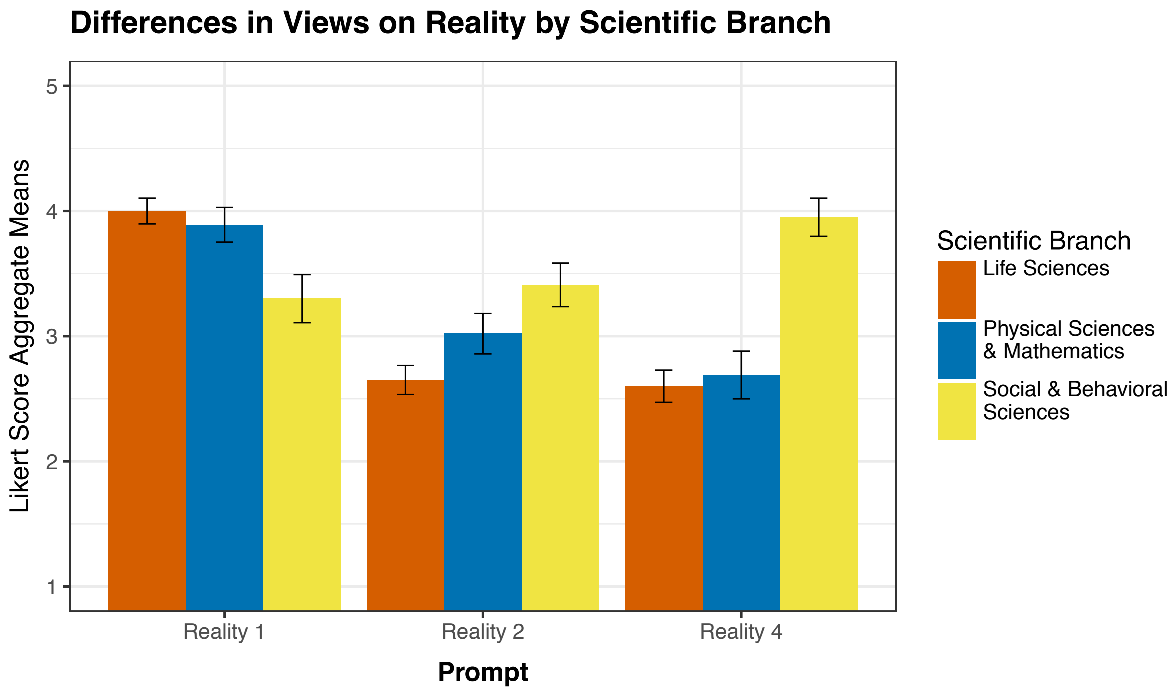

Figure 1.[15] Summary of means by scientific branch for the Reality 1, 2, and 4 prompts

4.2 Methodology

Next we conducted follow-up analyses for Methodology 1 ("Scientific research (applied or basic) must be hypothesis driven"), Methodology 2 ("In my disciplinary research, I employ primarily quantitative methods"), Methodology 3 ("In my disciplinary research, I employ primarily qualitative methods"), and Methodology 4 ("In my disciplinary research, I employ primarily experimental methods"). Beginning with Methodology 1, life scientists and physical scientists and mathematicians were more willing to agree that science must be driven by

---

[15] All figures use the colorblind-friendly palette developed by Okabe & Ito (2002).

hypotheses than were social and behavioral scientists ($U = 4949$, $p = 0.002$, $r = 0.22$ and $U = 2627$, $p = 0.009$, $r = 0.23$ respectively).[16] Again, Life Sciences and Physical Sciences & Mathematics were statistically indistinguishable (Figure 2).

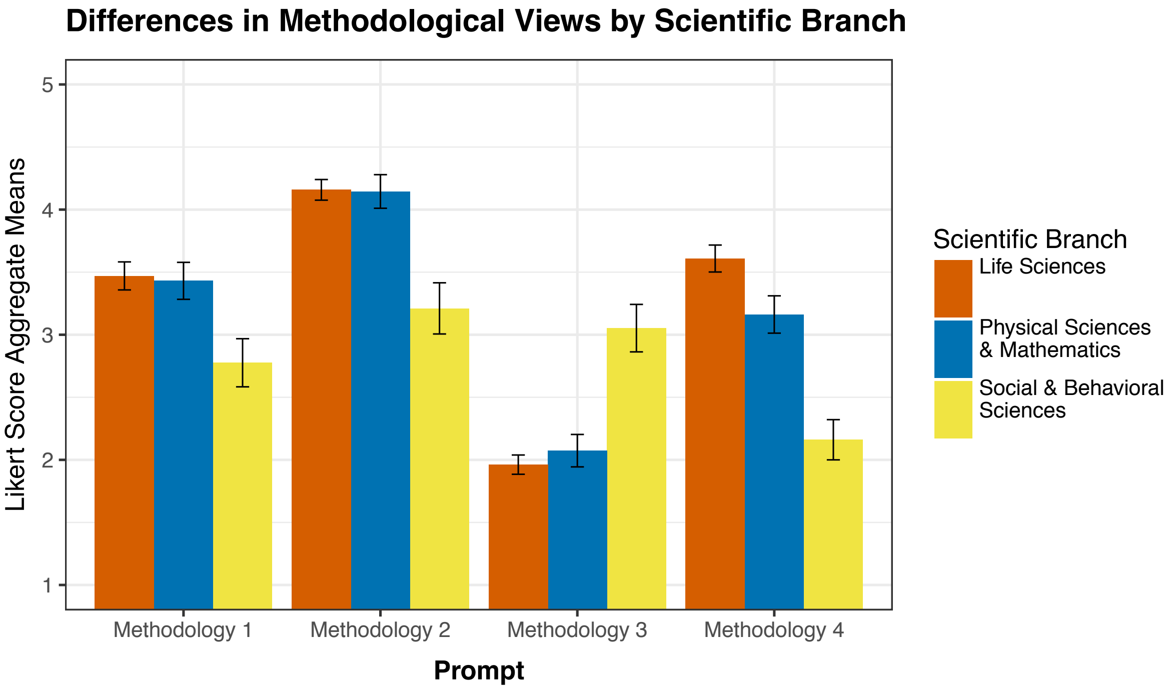

Figure 2. Summary of means by scientific branch for the prompts Methodology 1-4.

Methodology 2 and 3 concern quantitative and qualitative methods, respectively, and so can be discussed together. Both exhibit the previously observed pattern, where there is no significant difference between Life Sciences and Physical Sciences & Mathematics but both

---

[16] Recent qualitative analysis of Toolbox dialogues supports this point (Shannon M Donovan, O'Rourke, and Looney 2015).

differed from Social & Behavioral Sciences. Specifically, researchers in the Social & Behavioral Sciences are less inclined than researchers in the other two branches to employ quantitative methods ($U = 5052$, $p = 1.10 \times 10^{-04}$, $r = 0.28$ and $U = 2648.5$, $p = 3.99 \times 10^{-04}$, $r = 0.31$) and more inclined to employ qualitative methods ($U = 2109.5$, $p = 3.89 \times 10^{-07}$, $r = 0.36$ and $U = 1170$, $p = 5.91 \times 10^{-05}$, $r = 0.35$). Methodology 4 is the only prompt where all three branches are significantly different from one another. Life scientists were most likely to report employing experimental methods, followed by the physical scientists and mathematicians, and the social and behavioral scientists were the least likely to rely on experimentation.

4.3 Confirmation and Values

Finally, we consider the Confirmation and Values modules. Since only one prompt in each module was identified by the Kruskal–Wallis tests as requiring follow-up analysis, we present them together. As Figure 3 shows, the pattern of results is similar for these two prompts. Confirmation 3 (“Validation of evidence requires replication”) suggests that the demand for replication in the Social & Behavioral Sciences is significantly lower than in the Life Sciences ($U = 5309.5$, $p = 1.42 \times 10^{-05}$, $r = 0.31$) or Physical Sciences & Mathematics ($U = 2656.5$, $p = 0.002$, $r = 0.27$). With respect to Values 3 (“Value-neutral scientific research is possible”), social and behavioral scientists were less willing to endorse the possibility of value-neutral science than life scientists ($U = 4639$, $p = 0.002$, $r = 0.23$) and physical scientists and mathematicians ($U = 2586.5$, $p = 0.006$, $r = 0.24$). Future research on this topic is particularly recommended. In both cases, Life Sciences and Physical Sciences & Mathematics were statistically indistinguishable.

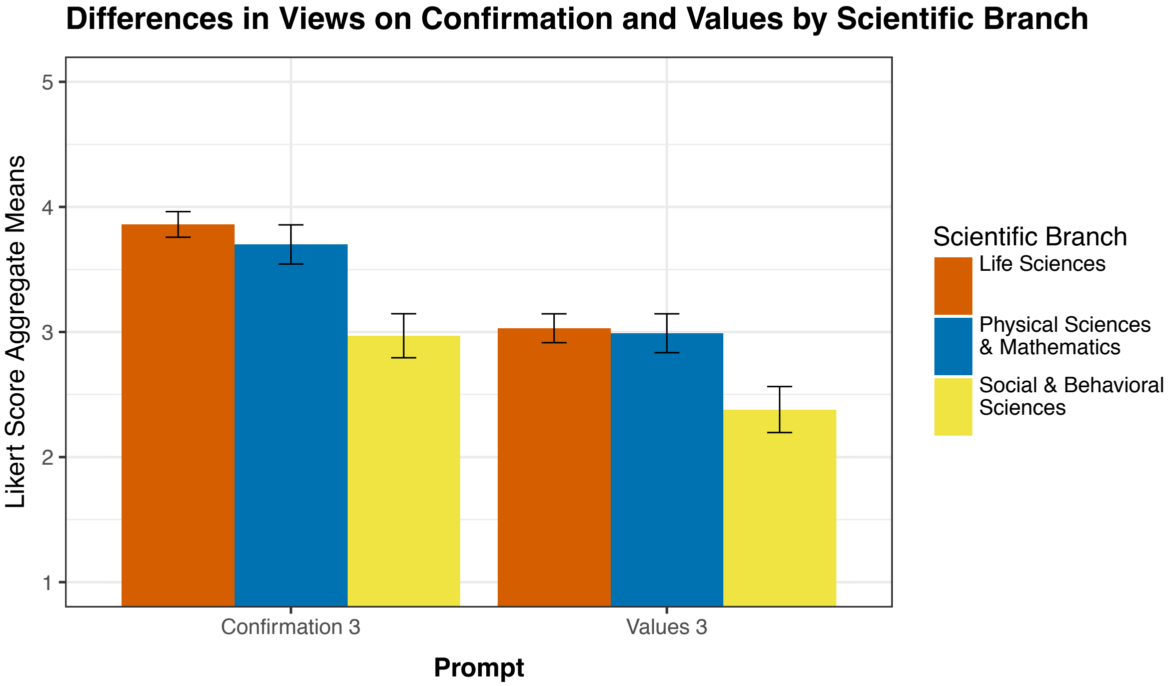

Figure 3. Summary of means by scientific branch for the Confirmation 3 and Values 3 prompts.

5 Discussion

In this section, we dig more deeply into our results by first addressing a concern about how we interpret participants' responses. We then turn to the project of distinguishing the natural and the social sciences. Finally, we consider the implications these results have for the ongoing conversation within philosophy of science about interdisciplinary integration.

5.1 Addressing a methodological worry

As we adumbrated above, our strategy has been to probe for philosophical differences among scientific branches by examining the philosophical commitments or attitudes of scientists who belong to these branches, as indicated by their responses of agreement and disagreement with Toolbox prompts. For nine prompts, we have found reason to think that scientists in different branches tend to respond differently. We might be inclined to draw a number of immediate conclusions from these findings. For example, we might be inclined to say that social and behavioral scientists *disagree* with life scientists, physical scientists, and mathematicians about whether value-neutral scientific research is possible. However, for reasons raised by an anonymous referee and mentioned in note 11, we need to proceed slowly.

For our four findings that involve reflexive prompts (viz., Methodology 2, 3, 4, and Reality 4), respondents are expressing attitudes about propositions that involve reference to themselves. In these cases, we cannot say that participants disagreed since the propositions to which they are responding differ. Nevertheless, we can note the differences. Social and behavioral scientists differ, for instance, to the extent to which they "employ primarily qualitative methods" in their disciplinary research. The remaining five prompts concern science understood more broadly, and so it is possible that representatives of the scientific branches disagree with one another about these, *so long as they are interpreting the propositions in the same way,* e.g., as claims about science in general. However, it might be that the participants are responding to the prompts in a way that reflects their particular scientific field, e.g., social science or natural science. So, a social scientist would interpret

Values 3 as a claim about value-neutral *social* science, whereas a physical scientist would interpret it as a claim about value-neutral *natural* science. In this case, the propositions differ, and so we could not conclude that different responses indicate disagreement.

We acknowledge that for the five non-reflexive prompts that support branch-level differences, our data do not rule out the possibility that participants are interpreting them differently. We are, however, reasonably confident that the participants interpreted these prompts in a non-localized and more general way, for three reasons. First, before the Toolbox workshop, participants were asked to read (Eigenbrode et al. 2007), which places the Toolbox approach, and thereby the workshop, in the context of a wide range of possible scientific contributions to interdisciplinary research. Second, prior to responding to the prompts, participants heard a preamble from their workshop facilitator that motivates the workshop as a response to communication issues that arise in interdisciplinary collaborations involving representatives of potentially any subset of scientific disciplines. Third, typical workshops involve representatives of several disciplines. Often the disciplines are epistemically distant (e.g., sociology and microbiology). The preamble aims to make it clear that the dialogue is intended to highlight similarities and differences across the disciplines represented in the room. Considered together, these factors encourage an orientation on the part of the participant toward science in general, an orientation that it is reasonable to think carries over to the interpretation of the non-reflexive prompts.

But, even if we are wrong and participants tended to interpret the non-reflexive prompts more locally, our principal conclusion about philosophical differences across scientific branches still holds. Although we are interested in whether or not the participants agree about the issues expressed by these prompts, our primary interest is in using our data to illuminate relationships among scientific branches. We maintain that this second broad conclusion follows whether the scientists involved tended to interpret the non-reflexive prompts in terms of science in general or their scientific fields in particular. To see this, assume first that the participating scientists tended to interpret the non-reflexive prompts in terms of science in general. Again, we found that a subset of these prompts gave rise to responses that tended to differ across scientific branches. On the current assumption, they are thus prompts give rise to cross-branch disagreements. This is a finding in need of an explanation. One reasonable explanation is that the disagreements stem from different views about science in general that reflect experiences of scientific branches that tend to differ in exactly these ways. On this assumption, what our results give us, then, is an indirect indication of what the Social & Behavioral Sciences, Life Sciences, and Physical Sciences & Mathematics are like.

Now suppose that the previous assumption is wrong. Assume that the participating scientists tended to interpret the non-reflexive prompts in terms of their respective fields. This could be more general, like social science or natural science, or more specific, like sociology or physics. In this case, there is an even more direct path from their Toolbox responses to conclusions about their scientific branches. If social scientists say *P* about social science and natural scientists say ~*P* about physical science, we should conclude that because social scientists are the best sources of insight into social science and natural scientists into physical science that *P* can be used to differentiate social science and natural science. Thus, either way, we are able to infer conclusions about their scientific branches from their Toolbox responses, which is the key conclusion we wish to draw from our data and the conclusion that will inform the rest of this Discussion section.

5.2 A closer look

There are a number of points to make about our results. First, they revealed little difference between Life Sciences and Physical Sciences & Mathematics. There was a statistically significant difference between them on only one prompt, Methodology 4. Otherwise, the two branches were indistinguishable. Only Social & Behavioral Sciences could be distinguished from one or both of the other scientific branches. On the assumption that Life Sciences and Physical Sciences & Mathematics qualify as natural sciences, differences like those reported here could help confirm philosophical accounts of what distinguishes the natural sciences from the social sciences. Consider Hacking's account, introduced above. According to this account, the social sciences differ from the natural sciences because the classificatory practices of the social sciences interact with the world differently than the classificatory practices of the natural sciences. If this is true, then social scientists would have greater exposure than natural scientists to the makings of a prima facie consideration against one of the pillars of scientific realism—the metaphysical claim that the sciences investigate a mind-independent world. As such, it wouldn't be too surprising to find social scientists more inclined toward antirealist attitudes about science than natural scientists. Our results with respect to Reality 1, 2, and 4 correspond to this expectation, thereby lending credence to Hacking's theory.

Further, our results provide some reason to think that Life Sciences and Physical Sciences & Mathematics are similar with respect to replication, value-neutrality, the role of hypotheses, and quantitative methods. Individually, these results are what we would expect—that both would be more inclined to use quantitative rather than qualitative methods, for example—but taken together, they contribute to an increasingly detailed map of philosophical common ground they share with one another but not with the Social & Behavioral Sciences, and this helps motivate philosophical efforts to distinguish natural and social sciences. Further research is warranted to fill out this map, as well as explore where there might be philosophical differences between the two branches (e.g., the role of experimentation).

A second point to make about our results stems from our finding that social and behavioral scientists were less willing to agree that value-neutral scientific research is possible (Values 3, $M = 2.38$, $S = 1.40$) than life scientists ($M = 3.03$, $SD = 1.29$) and physical scientists and mathematicians ($M = 2.99$, $SD = 1.30$). This difference might be due to a greater appreciation by scientists in the Social & Behavioral Sciences than in the other branches for the roles that values play in science—especially non-epistemic values. Such a hypothesis is a ripe topic for future research.[17] This is not to say that there was widespread endorsement for the possibility of value-neutral science in the other branches; the means for both branches were very close to the midpoint of the scale.[18] Nevertheless, as Steel et al. (2017) argue, the absence of an endorsement for value-neutral science suggests that philosophers of science should not assume, in building their normative theories about science or in developing their normative interventions for improving science, that scientists endorse the "value-free ideal," according to which good science is science that is not influenced by non-epistemic human values (Douglas 2009).

Furthermore, on the reasonable assumption that many scientists understand value neutrality to be one form that scientific objectivity can take, our results suggest that further

---

[17] For related empirical work on this topic, see Reiners et al. (2013), who focus on values in ecology.

[18] If the participants do not know how to respond to the prompts or are uncertain, they can select the response option "Don't Know". Each module contained an additional prompt focusing on the similarity of the group's views in that module (which we've omitted from analysis for this paper). For these prompts, 80 to 132 participants responded, "Don't Know." Additionally, for the Reductionism 1 prompt, 40 participants also responded, "Don't Know." The availability of this option and its regular use increases our confidence that means near the midpoint can be interpreted as "neither agree nor disagree."

empirical study is warranted to investigate if scientists' responses to issues related to scientific objectivity are mediated by their views on value-neutral science. This suggestion is motivated, in part, by Reiss and Sprenger (2016), who contend that:

> [m]any central debates in the philosophy of science have, in one way or another, to do with objectivity: confirmation and the problem of induction; theory choice and scientific change; realism; scientific explanation; experimentation; measurement and quantification; evidence and the foundations of statistics; evidence-based science; feminism and values in science.

For example, as an initial conjecture, it might be that scientists who reject value-neutrality would be less likely to require replication for confirmation. After all, rejecting value-neutrality suggests an openness to methods for evaluating scientific research that are sensitive to value differences, and such methods could permit variation in acceptable results that is inconsistent with the replication standard.

We note two caveats. First, our findings are preliminary and are based on a data set too small to provide resolution at the level of individual scientific *disciplines*. Given this limitation, it remains possible that the philosophical views of scientists correlate with scientific disciplines and we missed this, detecting only occasional correlations between branches and philosophical views. Second, it should be noted that the prompts in the STEM instrument are not validated using psychometric techniques. Their original purpose is to promote dialogue among collaborators, and to this end, they occasionally embed vague or ambiguous terminology designed to tease out unacknowledged differences in conceptualization among collaborators during a dialogue structured by their responses to these prompts. For example, Reality 4 includes the term "human construction," which allows for multiple interpretations. A sociologist might agree with this prompt, thinking the subject of their research (e.g., government, money) would not exist without humans, while an engineer might agree with it thinking that the subject of their research (e.g., bridges) are constructed by humans. Further research designed to build on these preliminary findings should seek to collect attitude data with prompts that are immune to this sort of concern.

5.3 Experimental data, philosophy of science, and interdisciplinary integration

We conclude our discussion by considering several implications our results have for integration in philosophy of science, and especially interdisciplinary integration. Interdisciplinarity is a "powerful trend in contemporary science" (Mäki 2016, 328) that involves the integration of distinct disciplinary explanations, data, standards, and methods (Brigandt 2013). This type of integration is also recognized as a key part of responses to the "Grand Challenges" facing humanity (Grandis and Efstathiou 2016).[19] As such, it is a topic that

---

[19] For example, an interdisciplinary account of the risks posed by geoengineering, such as solar radiation management projects, would require the synthesis, or integration, of inputs from the earth sciences, sociology, economics, and philosophy (Tuana et al. 2012). Although

has garnered increased attention from philosophers of science, especially those interested in interdisciplinarity, biology, or cognitive science (Bermúdez 2010; Brigandt 2010, 2013; Griesemer 2013; Grune-Yanoff 2016; Holbrook 2013; Leonelli 2013; O'Rourke, Crowley, and Gonnerman 2016; O'Rourke 2017).

Our empirical work positions us to make three contributions to the philosophical understanding of interdisciplinary integration in the sciences. Each can be seen as contributing to the philosophical task described by Brigandt: "to understand what [integration] involves, how integrative practices operate, how integrative accounts are formed, and what the challenges and limits to integration are" (2013, 461–62) . First, although philosophy can profit from the interaction of different disciplines in interdisciplinary research (Crowley, Gonnerman, and O'Rourke 2016), we have argued that philosophy can also be a source of conceptual differences that challenge interdisciplinary integration (Eigenbrode et al. 2007). Our work serves as an empirical contribution to the task of better understanding what this challenge looks like in particular collaborative contexts. Specifically, our findings reveal some of the contours of an emerging map of philosophical attitudes about science. While parts of the map remain obscured, our results indicate that social and behavioral scientists differ

---

interdisciplinary research does not require collaboration, it is often collaborative due to the need for a rich variety of inputs and a sensitivity to the range of disciplinary constraints on integration.

methodologically from biophysical scientists in several ways (Methodology 1-4). They are also inclined to be suspicious of the value-free ideal (Values 3), naïve realism about the external world (Realism 4), and objectivity as a regulative ideal (Realism 1 & 2). Biophysical scientists differ among themselves in terms of how reliant they are on experimentation (Confirmation 3). It is reasonable that these differences represent specific differences among branches of science. Interdisciplinary teams that include representatives of these scientific branches may need to negotiate these differences if their collaboration is going to function well (Lélé and Norgaard 2005). The methodological differences may appear unsurprising, they are not trivial in the context of interdisciplinary research teams trying to develop an integrated methodology. Navigating the differences can often become quite challenging (Eigenbrode et al. 2007; O'Rourke and Crowley 2013). The same can be said for the value-neutral ideal. It can't be determined *prima facie* which differences will be trivial or formidable in the life of an interdisciplinary research team.

A second point that our empirical work positions us to make about interdisciplinary integration relates to the challenge of integrating across the natural science/social science divide (Barthel and Seidl 2017; Lélé and Norgaard 2005). According to Rylance, interdisciplinarity that attempts to integrate "distant" disciplines is "more complex to undertake" (2015, 314). As a first approximation, it is reasonable to take disciplines in the Social & Behavioral Sciences branch to be more "distant" from those in, say, the Life Sciences than any two disciplines in the Social Sciences or in the Life Sciences would be from one another; this is suggested, after all, by their classification in different academic branches. Our results provide empirical support for this metric, highlighting a number of conceptual dimensions along which the social sciences differ from the natural sciences. When disciplines from these branches collaborate, the distance between how they conceptualize and practice science makes the common ground necessary to support interdisciplinary integration (Klein 2012) more difficult to come by, which could explain why these collaborations are more complex. In no small part, how one tends to conceptualize and practice science is grounded in one's philosophical attitudes about science, including one's core beliefs about the nature of science and the world and differences in operative values about scientific practice.

A third and more speculative point is that our results provide some reason to think that differences among the philosophical attitudes of scientists are not neatly partitioned into academic branches. There is considerable variability within branches—see the standard deviation values in Table 4 (Appendix)—suggesting that considerable differences in philosophical attitudes may exist within related disciplines. When we speak of interdisciplinary research projects, we talk about inter-*disciplinary* integration. But maybe *discipline* is best viewed as a proxy variable for a host of factors that individual researchers bring into the project. After all, scientists don't integrate disciplines when engaging in interdisciplinary integration; rather, they integrate methods, data, or perspectives as practiced, supplied, or championed by individuals. They seek individual agreement with decisions that combine perspectives in complex, often *sui generis* ways (Klein 2012). Differences in core beliefs and values, including those pertaining to science and its practice, may partition at the level of discipline in interesting ways, but it is also possible that in any collection of collaborators—even those that might seem like "unidisciplinary" collaborations—the differences in core beliefs and values among the individuals represent obstacles to meaningful integration. So, in short, by focusing on disciplines in interdisciplinary integrations, we may be looking in the wrong place—perhaps the really important differences do not aggregate at the disciplinary level.

6 Conclusion

Our examination of Toolbox data revealed differences in how scientists from different branches of science reacted to philosophical claims about aspects of scientific practice. In total, these results suggest that distinctions among branches of science are correlated with interesting philosophical differences. These differences concern attitudes toward realism, objectivity, replication, values, hypotheses, and methods. Since our results are preliminary, further experimental research in the philosophy of science is warranted to explore the scope and nature of these differences, as well as the extent to which these differences are responsible for deeper differences among ways of practicing science.

These results make two contributions to the burgeoning field of experimental philosophy of science. First, reflection on our results, and especially their bearing on recent work on scientific integration, provides an argument by example for the validity and effectiveness of experimental philosophy of science. These results help to reinforce philosophical conclusions about scientific integration derived through non-experimental means (e.g., the claim that integration across the natural/social divide can be especially challenging to achieve). In this way, experimental philosophy of science can provide a line of converging evidence in support of extant views in philosophy of science. Second, our results show that experimental philosophy of science can also suggest insights into science that may be harder to gain through non-experimental means, e.g., that philosophical differences across individuals may be more of an obstacle to interdisciplinary integration than cross-disciplinary differences.

Appendix

	Comparison	*U*	*p*	*r*
Reality 1	LS-PSM	4939	0.29	0.07
	LS-SBS*	4776.5	9.52×10^{-4}	0.24
	PSM-SBS*	2432.5	0.03	0.20
Reality 2	LS-PSM	3351	0.07	0.13
	LS-SBS*	2364.5	3.79×10^{-4}	0.26
	PSM-SBS	1469	0.08	0.16
Reality 4	LS-PSM	3409	0.69	0.03
	LS-SBS*	1698.5	2.70×10^{-9}	0.43
	PSM-SBS*	875	2.90×10^{-6}	0.42
Confirmation 3	LS-PSM	4888.5	0.54	0.04
	LS-SBS*	5309.5	1.42×10^{-5}	0.31

	PSM-SBS*	2656.5	0.002	0.27
Values 3	LS-PSM	4474.5	0.78	0.02
	LS-SBS*	4639	0.002	0.23
	PSM-SBS*	2586.5	0.006	0.24
Methodology 1	LS-PSM	4950	0.75	0.02
	LS-SBS*	4949	0.002	0.22
	PSM-SBS*	2627	0.009	0.23
Methodology 2	LS-PSM	4400.5	0.61	0.04
	LS-SBS*	5052	1.10×10^{-4}	0.28
	PSM-SBS*	2648.5	3.99×10^{-4}	0.31
Methodology 3	LS-PSM	4352	0.78	0.02
	LS-SBS*	2109.5	3.87×10^{-7}	0.36
	PSM-SBS*	1170	5.91×10^{-5}	0.35
Methodology 4	**LS-PSM***	5472.5	0.01	0.18
	LS-SBS*	5880	2.00×10^{-11}	0.47
	PSM-SBS*	2733.5	1.37×10^{-5}	0.38

Table 3. Summary of results for the Mann–Whitney–Wilcoxon (MWW) test for the nine prompts suggested for further analysis by the Kruskal–Wallis tests. Comparisons marked in bold and with asterisks show significant differences.

Prompt	Mean	Standard Deviation	Prompt	Mean	Standard Deviation
Motivation 1	3.67	1.08	Reality 1	3.81	1.26
Motivation 2	2.90	1.30	Reality 2	2.92	1.32
Motivation 3	2.86	1.29	Reality 3	3.11	1.24
Motivation 4	3.68	1.20	Reality 4	2.95	1.48
Methodology 1	3.31	1.35	Values 1	2.07	1.27
Methodology 2	3.95	1.21	Values 2	2.05	0.99
Methodology 3	2.23	1.16	Values 3	2.87	1.34
Methodology 4	3.18	1.35	Values 4	3.29	1.35
Methodology 5	2.97	1.19	Values 5	2.78	1.19
Confirmation 1	3.84	0.99	Reductionism 1	2.91	1.28
Confirmation 2	3.72	1.09	Reductionism 2	2.10	1.13
Confirmation 3	3.62	1.29	Reductionism 3	3.62	1.09
Confirmation 4	3.89	0.96	Reductionism 4	2.93	1.29
Confirmation 5	4.41	0.82	Reductionism 5	4.35	0.85

Table 4. Summary of means and standard deviation by prompt

References

Barthel, Roland, and Roman Seidl. 2017. “Interdisciplinary Collaboration between Natural and Social Sciences – Status and Trends Exemplified in Groundwater Research.” *PLOS ONE* 12 (1): 1–27. https://doi.org/10.1371/journal.pone.0170754.

Bechtel, William. 1993. “Integrating Sciences by Creating New Disciplines: The Case of Cell Biology.” *Biology and Philosophy* 8 (3): 277–99. https://doi.org/10.1007/BF00860430.

Bepress. 2014. “Disciplines: Introduction to Digital Commons Three-Tiered Taxonomy of Academic Disciplines.” 2014. http://digitalcommons.bepress.com/cgi/viewcontent.cgi?article=1008&context=reference.

Bermúdez, José Luis. 2010. *Cognitive Science: An Introduction to the Science of the Mind*. Cambridge: Cambridge University Press.

Bird, Alexander. 2008. “The Historical Turn in Philosophy of Science.” In *Routledge Companion to the Philosophy of Science*, edited by Stathis Psillos and Martin Curd, 67–77. New York: Routledge.

Bracken, L. J., and E. A. Oughton. 2006. “‘What Do You Mean?’ The Importance of Language in Developing Interdisciplinary Research.” *Transactions of the Institute of British Geographers*. https://doi.org/10.1111/j.1475-5661.2006.00218.x.

Brigandt, Ingo. 2010. “Beyond Reduction and Pluralism: Toward an Epistemology of Explanatory Integration in Biology.” *Erkenntnis* 73 (3): 295–311. https://doi.org/10.1007/s10670-010-9233-3.

———. 2013. “Integration in Biology: Philosophical Perspectives on the Dynamics of Interdisciplinarity.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4 Pt A): 461—465. https://doi.org/10.1016/j.shpsc.2013.09.009.

Carruthers, Peter, Stephen Stich, and Michael Siegal, eds. 2002. *The Cognitive Basis of*

Science. Cambridge: Cambridge University Press.

Crowley, Stephen J., Chad Gonnerman, and Michael O'Rourke. 2016. "Cross-Disciplinary Research as a Platform for Philosophical Research." *Journal of the American Philosophical Association*. https://doi.org/10.1017/apa.2016.16.

Darden, Lindley, and Nancy Maull. 1977. "Interfield Theories." *Philosophy of Science* 44 (1): 43–64. https://doi.org/10.1086/288723.

Donovan, S. M., M. O'Rourke, and C. Looney. 2015. "Your Hypothesis or Mine? Terminological and Conceptual Variation Across Disciplines." *SAGE Open*. https://doi.org/10.1177/2158244015586237.

Donovan, Shannon M, Michael O'Rourke, and Chris Looney. 2015. "Your Hypothesis or Mine? Terminological and Conceptual Variation Across Disciplines." *SAGE Open* 5 (2): 2158244015586237. https://doi.org/10.1177/2158244015586237.

Douglas, Heather. 2009. *Science and the Value-Free Ideal*. Pittsburgh, PA: University of Pittsburgh Press.

Eigenbrode, Sanford D., Michael O'Rourke, J. D. Wulfhurst, David M. Althoff, Caren S. Goldberg, KAYLANI MERRILL, WAYDE MORSE, et al. 2007. "Employing Philosophical Dialogue in Collaborative Science." *BioScience*. https://doi.org/10.1641/B570109.

Faust, David, and Paul E. Meehl. 2002. “Using Meta-Scientific Studies to Clarify or Resolve Questions in the Philosophy and History of Science.” *Philosophy of Science* 69 (S3): S185–96. https://doi.org/10.1086/341845.

Feist, Gregory J., and Michael E Gorman, eds. 2013. *Handbook of the Psychology of Science*. New York: Springer.

Fiest, Gregoriy J. 2006. *The Psychology of Science and the Origins of the Scientific Mind*. New Haven, CT: Yale University Press.

Gerson, Elihu M. 2013. “Integration of Specialties: An Institutional and Organizational View.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4 Pt A): 515—524. https://doi.org/10.1016/j.shpsc.2012.10.002.

Grandis, Giovanni De, and Sophia Efstathiou. 2016. “Introduction—Grand Challenges and Small Steps.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 56: 39–47. https://doi.org/https://doi.org/10.1016/j.shpsc.2015.11.009.

Grantham, Todd A. 2004. “Conceptualizing the (Dis)Unity of Science.” *Philosophy of Science* 71 (2): 133–55. https://doi.org/10.1086/383008.

Griesemer, James. 2013. “Integration of Approaches in David Wake’s Model-Taxon Research Platform for Evolutionary Morphology.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4 Pt A): 525—536.

https://doi.org/10.1016/j.shpsc.2013.03.021.

Griffiths, Paul, and Karola Stotz. 2008. “Experimental Philosophy of Science.” *Philosophy Compass* 3 (3): 507–21. https://doi.org/10.1111/j.1747-9991.2008.00133.x.

Griffitsh, Paul, Edouard Machery, and Stefan Linquist. 2009. “The Vernacular Concept of Innateness.” *Mind & Language* 24 (5): 605–30. https://doi.org/10.1111/j.1468-0017.2009.01376.x.

Grover, S. C. 1981. *Toward a Psychology of the Scientist: Implications of Psychology for Contemporary Philosophy of Science*. Washington, D.C.: University Press of America.

Grune-Yanoff, Till. 2016. “Interdisciplinary Success without Integration.” *European Journal for Philosophy of Science* 6 (3): 343–60. https://doi.org/10.1007/s13194-016-0139-z.

Hacking, Ian. 1996. “The Disunities of the Sciences.” In *The Disunity of Science*, edited by P. Galison and D. Stump. Stanford, CA: Stanford University Press.

Hartner, Daniel F. 2013. “Conceptual Analysis as Armchair Psychology: In Defense of Methodological Naturalism.” *Philosophical Studies* 165 (3): 921–37. https://doi.org/10.1007/s11098-012-9981-9.

Holbrook, J Britt. 2013. “What Is Interdisciplinary Communication? Reflections on the Very Idea of Disciplinary Integration.” *Synthese* 190 (11): 1865–79.

Holm, Sture. 1979. “A Simple Sequentially Rejective Multiple Test Procedure.”

Scandinavian Journal of Statistics 6 (2): 65–70. http://www.jstor.org/stable/4615733.

Houts, Arthur C. 1989. “Contributions of the Psychology of Science to Metascience: A Call for Explorers.” In *Psychology of Science: Contributions to Metascience.*, 47–88. New York, NY, US: Cambridge University Press.

Klein, Julie Thompson. 2012. “Case Studies in Interdisciplinary Research.” Thousand Oaks, California: SAGE Publications, Inc. https://doi.org/10.4135/9781483349541.

Knobe, Joshua, Wesley Buckwalter, Shaun Nichols, Philip Robbins, Hagop Sarkissian, and Tamler Sommers. 2012. “Experimental Philosophy.” *Annual Review of Psychology* 63 (1): 81–99. https://doi.org/10.1146/annurev-psych-120710-100350.

Knobe, Joshua, and Richard Samuels. 2013. “Thinking like a Scientist: Innateness as a Case Study.” *Cognition* 126 (1): 72–86. https://doi.org/https://doi.org/10.1016/j.cognition.2012.09.003.

Lélé, Sharachchandra, and Richard B Norgaard. 2005. “Practicing Interdisciplinarity.” *BioScience* 55 (11): 967–75. http://dx.doi.org/10.1641/0006-3568(2005)055[0967:PI]2.0.CO.

Leonelli, Sabina. 2013. “Integrating Data to Acquire New Knowledge: Three Modes of Integration in Plant Science.” *Studies in History and Philosophy of Biol & Biomed Sci* 44 (4): 503–14.

Linquist, Stefan, Edouard Machery, Paul E. Griffiths, and Karola Stotz. 2011. “Exploring the Folkbiological Conception of Human Nature.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 366 (1563): 444–53. https://doi.org/10.1098/rstb.2010.0224.

Looney, Chris, Shannon Donovan, Michael O’Rourke, Stephen Crowley, Sanford Eigenbrode, Liela Rotschy, Nilsa Bosque-Pérez, and J. D. Wulfhorst. 2013. “Seeing through the Eyes of Collaborators: Using Toolbox Workshops to Enhance Cross-Disciplinary Communication.” In *Enhancing Communication & Collaboration in Interdisciplinary Research*, edited by Michael O’Rourke, Sanford Eigenbrode, and J. D. Wulfhorst. Thousand Oaks, CA: SAGE Publications.

Love, Alan C. 2008. “Explaining Evolutionary Innovations and Novelties: Criteria of Explanatory Adequacy and Epistemological Prerequisites.” *Philosophy of Science* 75 (5): 874–86. https://doi.org/10.1086/594531.

Machery, Edouard. 2016. “Experimental Philosophy of Science.” In *A Companion to Experimental Philosophy*, edited by Justin Sytsma and Wesley Buckwalter, 475–90. Oxford: Wiley-Blackwell.

———. 2017. *Philosophy within Its Proper Bounds*. Oxford: Oxford University Press.

Machery, Edouard, and Kara Cohen. 2012. “An Evidence-Based Study of the Evolutionary Behavioral Sciences.” *The British Journal for the Philosophy of Science* 63 (1): 177–

226. http://dx.doi.org/10.1093/bjps/axr029.

Mäki, Uskali. 2016. “Philosophy of Interdisciplinarity. What? Why? How?” *European Journal for Philosophy of Science* 6 (3): 327–42. https://doi.org/10.1007/s13194-016-0162-0.

Martínez, María Laura. 2009. “Ian Hacking’s Proposal for the Distinction between Natural and Social Sciences.” *Philosophy of the Social Sciences* 39 (2): 212–34. https://doi.org/10.1177/0048393108324304.

Mitchell, Sandra D., Lorraine Daston, Gerd Gigerenzer, Neven Sesardic, and Peter B. Sloep. 1997. “The Whys and Hows of Interdisciplinarity.” In *Human by Nature: Between Biology and the Social Sciences*, edited by Peter Weingart, Sandra D. Mitchell, Peter J. Richerson, and Sabine Maasen, 103–150. Mahwah, NJ: Lawrence Erlbaum.

Nagatsu, Michiru. 2013. “Experimental Philosophy of Economics.” *Economics and Philosophy* 29 (2): 263–276. https://doi.org/10.1017/S0266267113000229.

Nagel, Ernst. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace, and World.

O’Rourke, Michael. 2017. “Comparing Methods for Cross-Disciplinary Research.” In *Oxford Handbook of Interdisciplinarity*, edited by Robert Frodeman, Julie Thompson Klein, and Roberto Carlos Dos Santos Pacheco, 2nd ed., 276–90. Oxford: Oxford University Press.

O'Rourke, Michael, Stephen Crowley, and Chad Gonnerman. 2016. "On the Nature of Cross-Disciplinary Integration: A Philosophical Framework." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, January.

O'Rourke, Michael, and Stephen J. Crowley. 2013. "Philosophical Intervention and Cross-Disciplinary Science: The Story of the Toolbox Project." *Synthese*. https://doi.org/10.1007/s11229-012-0175-y.

Okabe, Masataka, and Kei Ito. 2002. "Color Universal Design (CUD) - How to Make Figures and Presentations That Are Friendly to Colorblind People." 2002. http://jfly.iam.u-tokyo.ac.jp/color/.

Papineau, David. 2009. "Naturalism." *The Stanford Encyclopedia of Philosophy*.

Proctor, Robert W., and E.J. Capaldi, eds. 2012. *Psychology of Science: Implicit and Explicit Processes*. Oxford: Oxford University Press.

Reiners, Derek S., W. A. Reiners, and Jeffrey A. Lockwood. 2013. "The Relationship between Environmental Advocacy, Values, and Science: A Survey of Ecological Scientists' Attitudes." *Ecological Applications* 23 (5): 1226–42. https://doi.org/10.1890/12-1695.1.

Reiss, Julian, and Jan Sprenger. 2016. "Scientific Objectivity." *The Stanford Encyclopedia of Philosophy*.

Rylance, Rick. 2015. “Grant Giving: Global Funders to Focus on Interdisciplinarity.” *Nature* 525 (7569): 313–15. https://doi.org/10.1038/525313a.

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. https://doi.org/https://doi.org/10.1016/j.shpsa.2017.04.002.

Stotz, Karola. 2009a. “Experimental Philosophy of Biology: Notes from the Field.” *Studies in History and Philosophy of Science Part A* 40 (2): 233–37. https://doi.org/https://doi.org/10.1016/j.shpsa.2009.03.008.

———. 2009b. “Philosophy in the Trenches: From Naturalized to Experimental Philosophy (of Science).” *Studies in History and Philosophy of Science Part A* 40 (2): 225–26. https://doi.org/https://doi.org/10.1016/j.shpsa.2009.03.011.

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): 647–73. https://doi.org/https://doi.org/10.1016/j.shpsc.2004.09.005.

Sytsma, Justin, and Jonathan Livengood. 2015. *The Theory and Practice of Experimental Philosophy*. Broadview Press.

Thagard, Paul. 1988. *Computational Philosophy of Science*. *Computational Philosophy of*

Science. Cambridge, MA, US: The MIT Press.

Tuana, Nancy, Ryan L Sriver, Toby Svoboda, Roman Olson, Peter J Irvine, Jacob Haqq-Misra, and Klaus Keller. 2012. "Towards Integrated Ethical and Scientific Analysis of Geoengineering: A Research Agenda." *Ethics, Policy & Environment* 15 (2): 136–57. https://doi.org/10.1080/21550085.2012.685557.

Wagenknecht, Susann, Nancy J. Nersessian, and Hanne Andersen, eds. 2015. *Empirical Philosophy of Science: Introducing Qualitative Methods into Philosophy of Science*. New York: Springer.

Weinberg, Jonathan M, and Stephen Crowley. 2009. "The X-Phi(Les): Unusual Insights into the Nature of Inquiry." *Studies in History and Philosophy of Science Part A* 40 (2): 227–32. https://doi.org/https://doi.org/10.1016/j.shpsa.2009.03.016.