

Hypothesis Testing in Scientific Practice: An Empirical Study

Moti Mizrahi, Florida Institute of Technology

Abstract: It is generally accepted among philosophers of science that hypothesis testing (or confirmation) is a key methodological feature of science. As far as philosophical theories of confirmation are concerned, some emphasize the role of deduction in confirmation (e.g., the H-D method), whereas others emphasize the role of induction in confirmation (e.g., Bayesian theories of confirmation). The aim of this paper is to contribute to our understanding of scientific confirmation (or hypothesis testing) in scientific practice by taking an empirical approach. I propose that it would be illuminating to learn how practicing scientists describe their methods when they test hypotheses and/or theories. I use the tools of data science and corpus linguistics to study patterns of usage in a large corpus of scientific publications mined from the JSTOR database. Overall, the results of this empirical survey suggest that there is an emphasis on mostly the inductive aspects of confirmation in the life sciences and the social sciences, but not in the physical and the formal sciences. The results also point to interesting and significant differences between the scientific subjects within these disciplinary groups that are worth investigating in future studies.

Keywords: Bayesian confirmation; confirmation; corpus linguistics; data science; hypothesis testing; hypothetico-deductive method

1. Introduction

It is generally accepted among philosophers of science that hypothesis testing is a key methodological feature of science. As Andersen and Hepburn (2016) put it, “Among the

activities often identified as characteristic of science are systematic observation and experimentation, inductive and deductive reasoning, and the formation and *testing of hypotheses and theories*” (emphasis added). As Rosenberg (2000, p. 112) points out, however, “testing hypotheses is by no means an easily understood matter.” Arguably, as far as theories of hypothesis testing (or the logic of confirmation) in philosophy of science are concerned, there are roughly two different approaches: the first emphasizes the *deductive* aspects of confirmation, whereas the second emphasizes the *inductive* aspects of confirmation. Falling under the approach that emphasizes the role of deduction in confirmation is the theory of hypothesis testing (or logic of confirmation) known as the Hypothetico-Deductive (H-D) method. As Andersen and Hepburn (2016) put it:

The standard starting point for a non-inductive analysis of the logic of confirmation is known as the Hypothetico-Deductive (H-D) method. In its simplest form, the idea is that a theory, or more specifically a sentence of that theory which expresses some hypothesis, is confirmed by its true *consequences* (emphasis added).

Along the same lines, Crupi (2016) writes that, “The central idea of hypothetico-deductive (HD) confirmation can be roughly described as ‘deduction-in-reverse’: evidence is said to confirm a hypothesis in case the latter, while not *entailed* by the former, is able to *entail* it, with the help of suitable auxiliary hypotheses and assumptions” (emphasis added). And Johansson (2016, p. 47) sums up the H-D method as follows:

- Put forth a hypothesis.
- Infer an empirically testable claim from the hypothesis and eventual auxiliary assumptions.
- Determine the veracity of the empirically testable claim via experiment and observation.

- Depending on whether the empirical implications are true or false, determine whether the hypothesis is supported or falsified.

Accordingly, the H-D method is characterized by philosophers of science as *logical* (hence, the logic of confirmation), particularly *deductive* (hence, hypothetico-deductivism), because it involves *deducing consequences* from hypotheses (plus suitable assumptions and auxiliary hypotheses). As Andersen and Hepburn (2016) put it, “On the hypothetico-deductive account, scientists work to come up with hypotheses from which true *observational consequences* can be *deduced*” (emphasis added).¹

Falling under the approach that emphasizes the role of induction in confirmation are statistical theories of confirmation, such as Bayesian confirmation theories. As Andersen and Hepburn (2016) observe, “Work in statistics has been crucial for understanding how theories can be tested empirically, and in recent decades a huge literature has developed that attempts to recast confirmation in Bayesian terms.” Unlike theories of confirmation that emphasize the role of deduction, which rely on the deductive-logical notions of logical consequence (or entailment) and refutation to describe the relationships between evidence and hypotheses, theories of confirmation that emphasize the role of induction rely on non-deductive notions, such as “inductive strength” and “evidential support” (Crupi 2016).

For instance, in Bayesian epistemology, it is often assumed that rational agents have credences (or degrees of belief) that can vary in strength. Now, the standard Bayesian condition by which evidence e provides evidential support for hypothesis h is the following:

$$p(h | e) > p(h)$$

¹ Along the same lines, Potochnik et al. (2019) define the H-D method as follows: “a method of hypothesis-testing; an expectation is *deductively inferred* from a hypothesis and compared with an observation; violation of the expectation *deductively refutes* the hypothesis, while a match with the expectation non-deductively boosts support for the hypothesis” (emphasis added).

That is, if the probability of h given e is greater than the probability of h , then e lends strong evidential support to h when e is the case. As Reiss (2017, p. 56) puts it, “Direct evidence speaks in favour of the hypothesis by showing that what we’d *expect* to be the case were the hypothesis true is actually the case” (emphasis added). One well-known problem with this basic condition is known as the problem of old evidence (Howson 1984), which arises when e is known, and thus $p(e) = 1$. If $p(e) = 1$, then $p(h | e) = p(h)$. This is where a distinction between *accommodation* and *prediction* can be useful. Roughly speaking, an accommodation is an empirical consequence of a hypothesis or theory that is not novel, i.e., it is known at the time the hypothesis or theory is tested, whereas a prediction is novel insofar as it is unknown at the time the hypothesis or theory is tested (see, e.g., Maher 1988. Cf. Lange 2001). Accordingly, if e is a newly discovered phenomenon (a novel prediction), rather than old evidence that was already known (a mere accommodation), then the discovery of e is more surprising, and thus provides stronger inductive support for h .

Of course, this is by no means a definitive solution to the problem of old evidence in Bayesian confirmation theory. For an overview of the debate, see Barnes (2018). For present purposes, the important point is merely that there are, broadly speaking, two approaches to confirmation (or hypothesis testing): one that emphasizes the deductive elements of confirmation and another that emphasizes the inductive elements of confirmation. The “deductive” approach, of which the H-D method is an exemplar, relies on deductive-logical notions, such as logical consequence (or entailment) and refutation. The emphasis on deduction is evident from the use of logical terms and phrases, such as ‘entail’ and ‘deducing consequences’. The “inductive” approach, of which some version of Bayesian confirmation theory is an exemplar, relies on statistical or probabilistic notions, such as strong evidential support and prediction (or

expectation). The emphasis on induction is evident from the use of statistical or probabilistic terms, such as ‘expect’ and ‘strong support’.

It is important to note that the deductive and inductive elements of confirmation (or hypothesis testing) have been emphasized to a greater or lesser extent by various philosophers of science. According to Hoyningen-Huene (2008, pp. 167-168), from a historical point of view, philosophical thinking about science and its methods can be divided into four phases:

Phase I (from antiquity to the seventeenth century): “In this phase, the specificity of scientific knowledge was seen in its absolute certainty. There was an essential contrast between *episteme* (knowledge) and *doxa* (belief), and only *episteme* qualified as science. Its certainty was established by proof from evident axioms” (Hoyningen-Huene 2008, p. 167).

Phase II (from the seventeenth century into the nineteenth century): In this phase, “the means to establish certainty have been generalized to include inductive procedures as well” (Hoyningen-Huene 2008, p. 168).

Phase III (from the second half of the nineteenth century to the late twentieth century): In this phase, “Empirical knowledge produced by the scientific method(s) was now assessed to be fallible. However, a special status was still ascribed to it due to its distinctive mode of production” (Hoyningen-Huene 2008, p. 168).

Phase IV (from the last third of the twentieth century to the present): “In this phase, belief in the existence of scientific methods of the said kind has eroded. Historical and

philosophical studies have made it highly plausible that scientific methods with the characteristics as posited in the second and third phase do not exist” (Hoyningen-Huene 2008, p. 168).

In fact, some philosophers of science have devised theories of confirmation that aim to give roughly equal weight to the deductive and inductive elements of hypothesis testing (or at least, theories that do not emphasize deductive over inductive, or inductive over deductive, elements of confirmation). See, for example, Kuipers’ “hypothetico-probabilistic (HP-) method” (Kuipers 2009).

The aim of this paper is to contribute to our understanding of scientific confirmation (or hypothesis testing) in scientific practice by taking an empirical approach. I propose that it would be illuminating to learn how practicing scientists describe their methods when they test hypotheses and/or theories. Do practicing scientists describe hypothesis testing in mostly deductive or inductive terms (or both)? That is, do practicing scientists use deductive terms, such as ‘consequence’, ‘implication’, ‘entailment’, and ‘refutation’, when they talk about testing hypotheses and/or theories in their published works, thereby indicating that there is an emphasis on the deductive aspects of confirmation in scientific practice? Do practicing scientists use inductive terms, such as ‘prediction’, ‘forecast’, ‘expectation’, and ‘strong support’, when they talk about testing hypotheses and/or theories in their published works, thereby indicating that there is an emphasis on the inductive aspects of confirmation in scientific practice?

I propose that the tools of data science can help us shed some new light on these questions. By using the text mining, corpus analysis, and data visualization techniques of data science, we can study large corpora of scientific texts in order to uncover patterns of usage. Those patterns of usage, in turn, could shed new light on theory confirmation (or hypothesis

testing) in scientific practice because what scientists say and do in their research publications clearly falls under “scientific practice” or “scientific activity.” In that respect, this empirical study of hypothesis testing (or confirmation) in scientific practice should be of particular interest to philosophers of science who advocate for “a conscious and organized programme of detailed and systematic study of scientific practice that does not dispense with concerns about truth and rationality” (Society for Philosophy of Science in Practice 2006-2019). To the extent that there has been a “practice turn” in philosophy of science, as some claim (Soler et al. 2014), empirical methods should be particularly useful to philosophers of science who are interested in studying scientific practices. According to the mission statement of the Society for Philosophy of Science in Practice (SPSP), “Practice consists of organized or regulated activities aimed at the achievement of certain goals” (Society for Philosophy of Science in Practice 2006-2019). Empirical methods, such as those used in this empirical study, namely, the text mining, corpus analysis, and data visualization techniques of data science, seem to be well suited for studying scientific activities, such as testing hypotheses and publishing the results in scientific journals, which scientists engage in when they aim at producing scientific knowledge.²

In the next section (Section 2), I will describe in more detail the empirical methods I have used in this empirical study of hypothesis testing (or confirmation) in scientific practice. In Section 3, I will report the results of this empirical study. In Section 4, I will discuss the implications of the results of this empirical study as far as our understanding of confirmation (or hypothesis testing) in science is concerned. Overall, the results of this empirical survey suggest that there is an emphasis on mostly the inductive aspects of confirmation in the life sciences and

² For more on the application of the empirical methods of data science, such as text mining and corpus analysis, to philosophy of science, see Mizrahi (2013), (2016), and (2020). For a recent example of an application of survey and other empirical methodologies from the social sciences to philosophy of science, see Beebe and Dellsén (2020).

the social sciences, but not in the physical and the formal sciences. The results also point to interesting and significant differences between the scientific subjects within these disciplinary groups that are worth investigating in future studies.

2. Methods

As discussed in Section 1, the research question that guides this empirical study of hypothesis testing in scientific practice is this: Do practicing scientists describe hypothesis testing in mostly deductive or inductive terms (or both)? More specifically:

(Q1) Do practicing scientists use deductive terms, such as ‘consequence’, ‘implication’, ‘entailment’, or ‘refutation’, when they talk about testing hypotheses and/or theories in scientific publications?

(Q2) Do practicing scientists use inductive terms, such as ‘prediction’, ‘forecast’, ‘expectation’, or ‘strong support’, when they talk about testing hypotheses and/or theories in scientific publications?

By adopting the methods of data science, I propose, we can find tentative answers to these questions empirically. The methods of data and text mining allow us to examine a large corpus of scientific texts (i.e., articles and book chapters published in scientific journals and books) in order to find out how practicing scientists talk about hypothesis testing in scientific publications. Such data can be mined from JSTOR Data for Research (www.jstor.org/dftr/). Researchers can use JSTOR Data for Research to create datasets, including metadata, n-grams, and word counts, for most of the articles and book chapters contained in the JSTOR database. JSTOR Data for Research is a particularly useful resource for the purposes of this empirical study because it provides an interface for creating datasets based on unique search queries and the associated

metadata for those search queries. By using this interface for constructing datasets, then, we can find out whether the aforementioned deductive and inductive confirmation terms appear in scientific publications and with what frequency relative to the total number of publications in a corpus.

The methods of data science allow us to overcome the limitations of relying on selected case studies from the history of science. For those case studies may or may not be representative of science as a whole. As Pitt (2001, p. 373) puts it, “if one starts with a case study, it is not clear where to go from there--for it is unreasonable to generalize from one case or even two or three.” Of course, empirical methodologies have limitations of their own. As far as the methods of text mining and corpus analysis are concerned, there are two major limitations. First, we can only study and analyze what is explicitly used in the corpus. For the purpose of this study, then, the corpus of scientific texts must contain explicit occurrences of deductive and/or inductive confirmation terms, e.g., instances of ‘consequence’, ‘prediction’, and the like, for us to be able to analyze means, proportions, and patterns of usage. It is reasonable to assume that there would be such explicit occurrences of deductive and/or inductive confirmation terms in scientific texts if hypothesis testing were a key methodological feature of science. Indeed, it would be quite surprising if, “Among the activities often identified as characteristic of science are systematic observation and experimentation, inductive and deductive reasoning, and the formation and *testing of hypotheses and theories*” (Andersen and Hepburn 2016), but confirmation terms were not explicitly used in scientific publications.

Second, as with many empirical methodologies, there may be some false positives and/or false negatives. When it comes to the methods of data science and corpus linguistics, false negatives could occur when we search for a specific term t in a corpus, but do not find it, even

though the corpus contains a synonym of t . For example, although unlikely, it is possible that our corpus of scientific texts contains no instances of ‘prediction’, and so a search for ‘prediction’ would return zero results, because scientists use a synonym for ‘prediction’ in all the publications that make up our corpus. On the other hand, false positives could occur when we find instances of a term t in our corpus, but those instances contain irrelevant uses of t . For the purpose of this empirical study, then, the corpus of scientific texts must contain not only explicit occurrences of deductive and/or inductive confirmation terms, e.g., instances of ‘consequence’, ‘prediction’, and the like, but also explicit occurrences of confirmation terms in the context of talk about hypotheses and/or theories. For example, instances of ‘prediction’ that are not about the prediction of a theory or theories would be considered false positives for the purposes of this empirical study.

Now, there are two things we can do to overcome the limitations of our empirical, data-driven approach. First, we can refine our search terms. As we have seen in Section 1, on the “deductive” approach to confirmation, hypotheses are tested by deducing from them logical consequences that can be observed (i.e., observational consequences). As Salmon (1970, p. 76) puts it (emphasis added):

H (hypothesis being tested)

A (auxiliary hypotheses)

I (initial conditions)

O (observational *consequence*)

Accordingly, we can use the term ‘consequence’ as an indicator that emphasis is given to deducing consequences from hypotheses or theories, i.e., to the deductive elements of

confirmation. This is a methodological assumption of this empirical study, namely, that deductive confirmation terms, such as ‘consequence’, are reliable indicators of an emphasis on the deductive elements of confirmation in scientific practice. In addition to ‘consequence’, we have seen in Section 1 that philosophers of science also use the terms ‘entailment’, ‘implication’, and ‘refutation’ to talk about the deductive aspects of confirmation. So we can use these terms as additional indicators that an emphasis is given to deductive implications or entailments from hypotheses or theories in scientific practice.

Likewise, as we have seen in Section 1, on the “inductive” approach to confirmation, a hypothesis is confirmed by the observation of a novel phenomenon that would be unlikely or improbable if the hypothesis in question were not true (i.e., observational predictions). As Salmon (1982, p. 49) puts it (emphasis added):

The hypothesis has a non-negligible prior probability.

If the hypothesis is true, then the observational *prediction* is very probably true. (If the hypothesis deductively implies the prediction, then this probability is 1.)

The observational *prediction* is true.

No other hypothesis is strongly confirmed by the truth of this observational *prediction*; that is, other hypotheses for which the same observational *prediction* is a confirming instance have lower prior probabilities.

Therefore, the hypothesis is confirmed.³

Accordingly, we can use the term ‘prediction’ as an indicator that emphasis is given to making observational predictions from hypotheses or theories, i.e., to the inductive elements of

³ See also Salmon (1976) on “Deductive” versus “Inductive” Archeology.

confirmation. Again, this is a methodological assumption of this empirical study, namely, that inductive confirmation terms, such as ‘prediction’, are reliable indicators of an emphasis on the inductive elements of confirmation in scientific practice. In addition to ‘prediction’, we have seen in Section 1 that philosophers of science also use the terms ‘expectation’, ‘forecast’, and ‘strong support’ to talk about the inductive aspects of confirmation (see Goodman 1983). So we can use these terms as additional indicators that an emphasis is given to inductive expectations or forecasts from hypotheses or theories in scientific practice.

If we include all of the aforementioned deductive and inductive confirmation terms in our search queries, we can be quite confident that we will not miss discussions of confirmation in scientific publications that are couched in synonymous terms, e.g., ‘expectation’ rather than ‘prediction’ or ‘entailment’ instead of ‘consequence’. This search methodology yields the search terms listed in Table 1. It is designed to minimize the number of false negatives.

Table 1. Search terms for approaches to confirmation that emphasize deduction or induction

Deductive confirmation terms	Inductive confirmation terms
consequence	prediction
implication	forecast
entailment	expectation
refutation	strong support

Second, we can make sure that our search methodology picks out instances of confirmation terms in the corpus that occur in the context of talk about hypotheses and/or theories. Since the aim of this paper is to find out how practicing scientists describe hypothesis testing in scientific practice, I have searched for confirmation terms in the context of talk about

hypotheses or theories by pairing the deductive and inductive confirmation terms listed in Table 1 with the scientific practice terms ‘hypothesis’ and ‘theory’. In practice, this means that I have searched for confirmation terms within ten words of the words ‘hypothesis’ or ‘theory’, e.g., (“hypothesis consequence”~10), (“hypothesis prediction”~10), (“theory implication”~10), (“theory forecast”~10), and so on, according to the following formulas:

(“deductive confirmation term scientific practice term”~10)

(“inductive confirmation term scientific practice term”~10)

It is important to note that, for proximity search to work properly in the JSTOR Data for Research’s dataset construction interface, the correct syntax must be used. In the case of proximity searches, such as the ones conducted for this empirical study, the syntax is (“term₁ term₂”~10), e.g., (“theory prediction”~10). Without the parentheses and quotation marks, a search query will yield search results that include text with more than ten words between term₁ and term₂. We would like to rule out such search results in order to avoid counting false positives. This syntax for proximity search, however, does not allow for wildcard searches using the asterisk symbol (*), e.g., (“predict* theory”~10). This search methodology is designed to minimize the number of false positives, i.e., instances of confirmation terms that are not about scientific hypotheses or theories, by ensuring that instances of the confirmation terms in text are anchored to the scientific practice terms ‘hypothesis’ or ‘theory’ (allowing for only ten words between a confirmation term, such as ‘consequence’, and a scientific practice term, such as ‘theory’).

To illustrate, here are a few examples of the search results that this search methodology picked out (emphasis added):

1. Life Sciences: “Prior to examining *predictions* from these *hypotheses*, it is valuable to examine whether variation exists in factors influencing group size and composition” (Treves and Chapman 1996, pp. 47-48).
2. Physical Sciences: “This *hypothesis* leads naturally to the further *consequence* that complete exhaustion will be approached asymptotically” (Russell 1919, p. 206).
3. Social Sciences: “From capital-dependence *theory* follows the *prediction* that firms going through a capital crisis will be particularly susceptible to introducing a CFO to their ranks” (Zorn 2004, p. 352).
4. Formal Sciences: “It is a trivial *consequence* of measure *theory*” (de Silva 2010, p. 918).

Apparently, then, there are instances of some of the deductive and inductive confirmation terms listed in Table 1 in scientific publications. Of course, we would like to know how frequent such instances are and whether terms that indicate one aspect of confirmation are more prevalent than terms that indicate another aspect of confirmation.

Contrary to the aforementioned examples, our search methodology will not count the following example of a false positive of ‘prediction’ as an occurrence of the inductive confirmation term ‘prediction’ in the corpus (emphasis added):

since the early 1970s, most associative *theories* of learning have incorporated the assumption that learning is driven by *prediction* error (Wills 2009, p. 96).

This is an example of a false positive of ‘prediction’ because the term ‘prediction’ is being used to talk about predictive errors made by subjects, not about the predictions of hypotheses or theories. Our search methodology will not count this instance of ‘prediction’ as an instance of the inductive confirmation term ‘prediction’ in scientific practice because there are more than ten

words between the scientific practice term ‘theory’ and the inductive confirmation term ‘prediction’.

Likewise, the following occurrence of the inductive confirmation term ‘strong support’ in the context of talk about hypotheses will be counted as a positive occurrence of an inductive confirmation term by our search methodology (emphasis added): “They found *strong support* for the *hypothesis*” (London 1992, p. 306). By contrast, the following occurrence of ‘strong’ and ‘support’ will not be so counted (emphasis added):

It could be that sharing norms and institutions can mediate the effects of POP on cultural evolution such that a small population with numerous and/or *strong* sharing norms and institutions is equivalent or even better in terms of its ability to retain beneficial inventions than a large population with few and/or weak sharing norms and institutions. If this is the case, then it is possible that the disagreement among the studies is the result of populations that *support* the *hypothesis* having fewer and/or weaker sharing norms and institutions than populations that do not *support* it (Collard et al. 2013, p. S396).

Since the terms ‘strong’ and ‘support’ are not collocated, our search methodology will not count this as an occurrence of an inductive confirmation term, which is exactly what we want our search methodology to do in this case. For what is being described as strong in this case is not the support for a hypothesis but rather sharing norms.

This search methodology is designed to test the following hypotheses about confirmation (or hypothesis testing) in scientific practice:

(H1) There is an emphasis on the deductive aspects of confirmation in scientific practice.

(H2) There is an emphasis on the inductive aspects of confirmation in scientific practice.

Assuming that the deductive and inductive confirmation terms listed in Table 1 are reliable indicators of emphasis on the deductive or inductive aspects of confirmation in scientific practice, respectively, these hypotheses would explain any observed proportions of confirmation terms in the corpus. That is, if there were an emphasis on the deductive elements of confirmation in scientific practice (H1), then we would expect to see more frequent occurrences of deductive confirmation terms than inductive confirmation terms in scientific publications. In other words, if practicing scientists describe confirmation (or hypothesis testing) in mostly deductive rather than inductive terms, then that would suggest that they emphasize the deductive aspects of confirmation in their published works. If the results of this empirical study were to bear out this expectation, then that would lend some empirical support to the hypothesis that there is an emphasis on the deductive aspects of confirmation in scientific practice.

On the other hand, if there were an emphasis on the inductive elements of confirmation in scientific practice (H2), then we would expect to see more frequent occurrences of inductive confirmation terms than deductive confirmation terms in scientific publications. In other words, if practicing scientists describe confirmation (or hypothesis testing) in mostly inductive rather than deductive terms, then that would suggest that they emphasize the inductive aspects of confirmation in their published works. If the results of this empirical study were to bear out this expectation, then that would lend some empirical support to the hypothesis that there is an emphasis on the inductive aspects of confirmation in scientific practice.

In that respect, it is important to note that, just like any other empirical study, the results of this empirical study are not to be interpreted as conclusive evidence for or against any hypothesis about confirmation (or hypothesis testing) in scientific practice. Nor are the methods used in this empirical study the only (or even the best) methods to study how practicing scientists

describe confirmation (or hypothesis testing) in scientific practice. Rather, they are supposed to add to our understanding of confirmation (or hypothesis testing) in science. Other studies, which make use of different empirical methods, such as survey procedures, can do the same (see, e.g., Beebe and Dellsén (2020)). In that sense, the results of this empirical study should be construed as tentative in the same sense that scientific conclusions are provisional (Marcum 2008).

The JSTOR database allows for searches by subject, such as Biological Sciences, Physics, and Sociology. In order to have a large and diverse sample that could be representative of science as a whole, I have conducted my searches on data mined from the Biological Sciences, Botany & Plant Sciences, Ecology & Evolutionary Biology, Astronomy, Chemistry, Physics, Anthropology, Psychology, Sociology, Computer Science, Mathematics, and Statistics subjects in the JSTOR database. That way, my datasets contain representative disciplines from the life sciences (namely, Biological Sciences, Botany & Plant Sciences, and Ecology & Evolutionary Biology), representative disciplines from the physical sciences (namely, Astronomy, Chemistry, and Physics), representative disciplines from the social sciences (namely, Anthropology, Psychology, and Sociology), and representative disciplines from the formal sciences (namely, Computer Science, Mathematics, and Statistics). All the searches for this empirical study were verified on March 9, 2020.

3. Results

Before we can see the results of the searches for the deductive and inductive confirmation terms listed in Table 1, it is useful to see how frequently practicing scientists use the scientific practice terms ‘hypothesis’ and/or ‘theory’ in their published work. This will then provide us with the base rates for our searches of the deductive and inductive confirmation terms listed in Table 1.

That is, we would like to know how many of the instances of ‘hypothesis’ and/or ‘theory’ in scientific publications are associated with the search terms for deductive or inductive confirmation listed in Table 1. The results of these searches are listed in Table 2.⁴

Table 2. Proportions of publications that contain the scientific practice terms ‘hypothesis’ and/or ‘theory’ in the total number of publications in the JSTOR database by subject (Source: JSTOR Data for Research)

	total	hypothesis	hypothesis/total	theory	theory/total
Biological Sciences	1322419	265538	0.20	235011	0.17
Botany & Plant Sciences	456408	62164	0.13	36494	0.07
Ecology & Evolutionary Biology	356294	96585	0.27	93561	0.26
Astronomy	18337	1988	0.10	3722	0.20
Chemistry	781	104	0.13	247	0.31
Physics	5584	1111	0.19	3274	0.58
Anthropology	335332	27564	0.08	85442	0.25
Psychology	90919	23921	0.26	44753	0.49
Sociology	717056	69076	0.09	256610	0.35
Computer Science	16793	1922	0.11	8725	0.51
Mathematics	367525	59318	0.16	184168	0.50
Statistics	135454	34719	0.25	81824	0.60

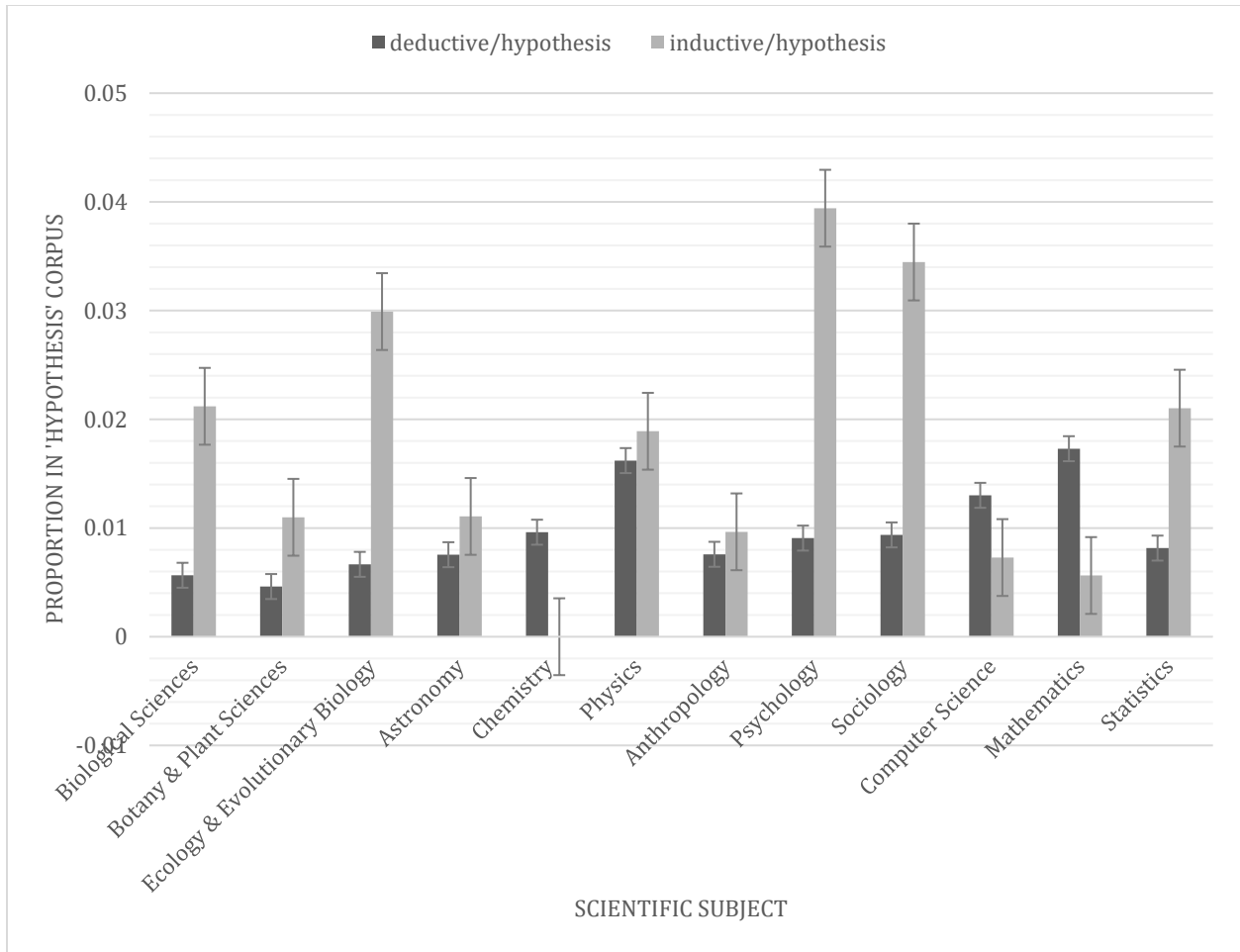
⁴ It is worth noting that the JSTOR database does not contain the same number of publications in each subject. In other words, some subjects (e.g., Biological Sciences) contain more publications than other subjects (e.g., Chemistry) in the JSTOR database. This should not make a significant difference to the results of this empirical study because the comparisons made are between *proportions* rather than raw numbers of publications from each subject.

Now that we have our prior probabilities of scientific publications that contain the scientific practice terms, namely, ‘hypothesis’ and/or ‘theory’, for each subject, we can look at how frequently these terms occur in conjunction with (i.e., within ten words of) the search terms for deductive and inductive confirmation listed in Table 1. That is, we would like to know how frequently deductive and inductive confirmation terms are invoked in the context of talk about hypotheses or theories. Accordingly, proportions will be calculated by taking the search results for each confirmation term and dividing it by the number of publications that contain hypothesis talk or theory talk, respectively. For example, 20% of Biological Sciences publications contain hypothesis talk. Now, of those publications, how many contain occurrences of the confirmation terms listed in Table 1? These results will be reported next.

3.1. Confirmation terms in the context of ‘hypothesis’ talk

Let us begin with the scientific practice term ‘hypothesis’ and the deductive and inductive confirmation terms listed in Table 1. The results of these searches are summarized in Figure 1.

Figure 1. Proportions of publications that contain deductive versus inductive confirmation terms within ten words of the scientific practice term ‘hypothesis’ in the total number of publications by subject (Source: JSTOR Data for Research)



As we can see from Figure 1, inductive confirmation terms are generally more frequent than deductive confirmation terms across scientific publications that contain discussion of hypotheses, with the exception of Chemistry, Computer Science, and Mathematics. I have conducted z-tests for proportion in order to find out whether the differences between the proportions of inductive and deductive confirmation terms are statistically significant within scientific subjects in the ‘hypothesis’ corpus.

In Biological Sciences, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 49.21$, $p = 0.00$, two-sided). In Botany & Plant Sciences, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically

significant ($z = 12.76$, $p = 0.00$, two-sided). In Ecology & Evolutionary Biology, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 38.14$, $p = 0.00$, two-sided). These results suggest that inductive confirmation terms are invoked significantly more often than deductive confirmation terms in life science publications that contain discussions of hypotheses.

In Astronomy, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is not statistically significant ($z = 1.15$, $p = 0.24$, two-sided). In Chemistry, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is not statistically significant ($z = 1.002$, $p = 0.31$, two-sided). In Physics, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is not statistically significant ($z = 0.48$, $p = 0.62$, two-sided). These results suggest that there is no significant difference between the frequency with which deductive and inductive confirmation terms are invoked in physical science publications that contain discussions of hypotheses.

In Anthropology, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 2.62$, $p = 0.00$, two-sided). In Psychology, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 21.57$, $p = 0.00$, two-sided). In Sociology, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 31.86$, $p = 0.00$, two-sided). These results suggest that inductive confirmation terms are invoked significantly more often than deductive confirmation terms in social science publications that contain discussions of hypotheses.

In Computer Science, the difference between the proportion of deductive confirmation terms and the proportion of inductive confirmation terms is not statistically significant ($z = -1.77$, $p = 0.07$, two-sided). This result suggests that there is no significant difference between the frequency with which deductive and inductive confirmation terms are invoked in Computer Science publications that contain discussions of hypotheses. In Mathematics, the difference between the proportion of deductive confirmation terms and the proportion of inductive confirmation terms is statistically significant ($z = -18.87$, $p = 0.00$, two-sided). This result suggests that deductive confirmation terms are invoked significantly more often than inductive confirmation terms in Mathematics publications that contain discussions of hypotheses. In Statistics, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 14.14$, $p = 0.00$, two-sided). This result suggests that inductive confirmation terms are invoked significantly more often than deductive confirmation terms in Statistics publications that contain discussions of hypotheses.

To check that the search methodology described in Section 2 returns genuine instances of the phenomenon in question (namely, instances of deductive and/or inductive confirmation terms in the context of talk about hypotheses), I have selected three search results from the ‘hypothesis’ corpus at random (emphasis added):

1. Life Sciences: “The *hypothesis* that HIV-1 MA forms trimers to accommodate the long gp41 CT has *implications* for the structures of related viruses” (Tedbury et al. 2016, p. E188).
2. Physical Sciences: “we review direct and indirect *predictions* and tests of the PN Binary *Hypothesis*” (De Marco 2009, p. 317).

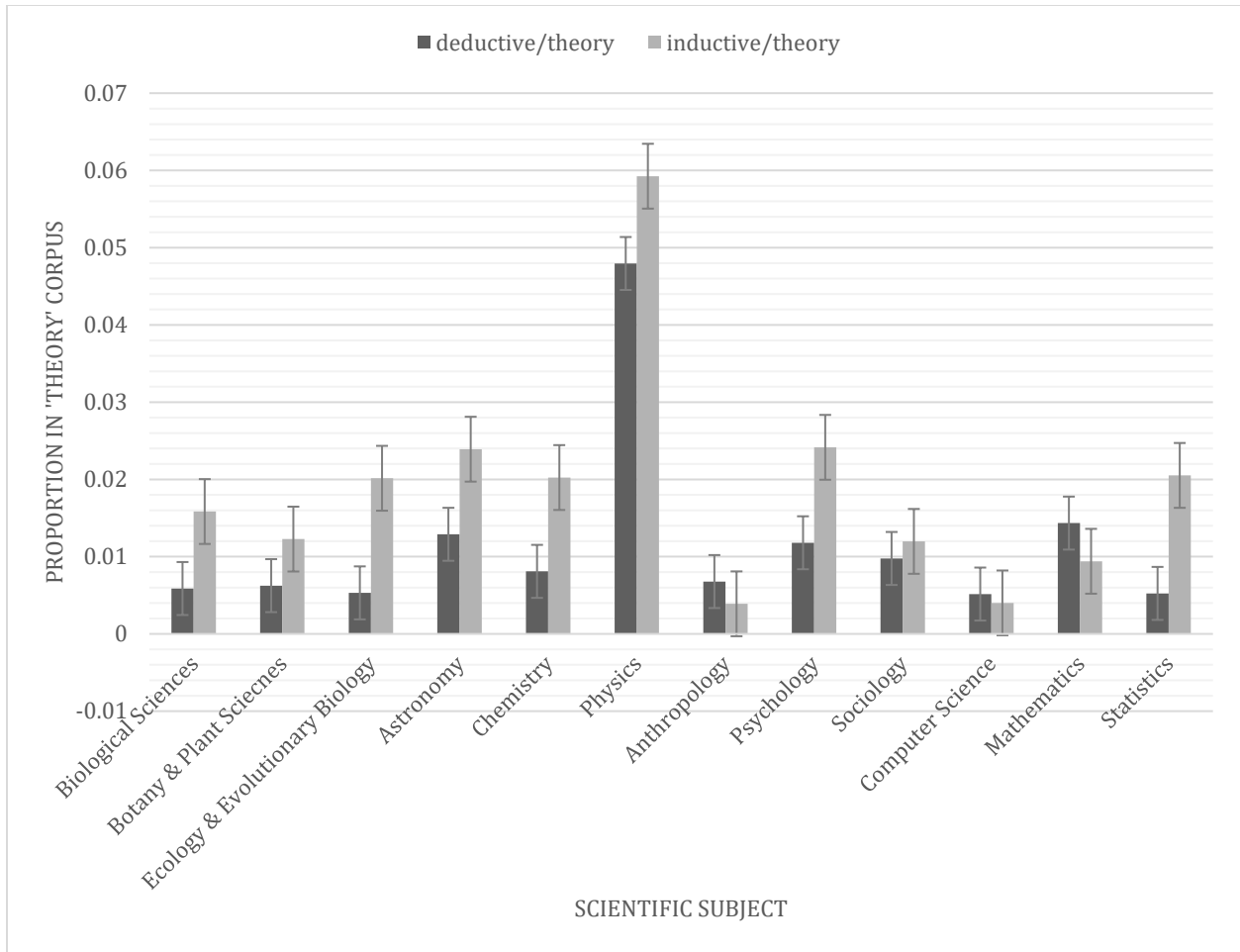
3. Social Sciences: “The other *expectation (hypothesis 2)*, i.e. that aggression and activity are not, or to a far lesser degree, related to depression, was only supported by the correlation between activity and depression” (Trijsburg et al. 1989, p. 197).
4. Formal Sciences: “such condition mismatch is a direct *consequence* of different biological *hypothesis* [sic] of interest” (Ruan and Yuan 2011, p. 1623).

These instances of deductive and/or inductive confirmation terms in research articles published in scientific journals also provide context to the statistical results reported above. They illustrate how practicing scientists use deductive and/or inductive confirmation terms when they talk about hypotheses in scholarly scientific practice.

3.2. Confirmation terms in the context of ‘theory’ talk

Now let us look at the results for the scientific practice term ‘theory’ and the deductive and inductive confirmation terms listed in Table 1. The results of these searches are summarized in Figure 2.

Figure 2. Proportions of publications that contain deductive versus inductive confirmation terms within ten words of the scientific practice term ‘theory’ in the total number of publications by subject (Source: JSTOR Data for Research)



As we can see from Figure 2, inductive confirmation terms are generally more frequent than deductive confirmation terms across scientific publications that contain discussion of theories as well, but now with the exception of Anthropology, Computer Science, and Mathematics. I have conducted z-tests for proportion in order to find out whether the differences between the proportions of inductive and deductive confirmation terms are statistically significant within scientific subjects in the ‘theory’ corpus.

In Biological Sciences, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 32.96$, $p = 0.00$, two-sided). In Botany & Plant Sciences, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically

significant ($z = 8.501$, $p = 0.00$, two-sided). In Ecology & Evolutionary Biology, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 28.62$, $p = 0.00$, two-sided). These results suggest that inductive confirmation terms are invoked significantly more often than deductive confirmation terms in life science publications that contain discussions of theories. As far as the life sciences are concerned, this is the same pattern we have observed in the ‘hypothesis’ corpus.

In Astronomy, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 3.53$, $p = 0.00$, two-sided). In Chemistry, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is not statistically significant ($z = 1.14$, $p = 0.25$, two-sided). In Physics, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 2.03$, $p = 0.04$, two-sided). These results suggest that inductive confirmation terms are invoked significantly more often than deductive confirmation terms in Astronomy and Physics, but not in Chemistry, publications that contain discussions of theories.

In Anthropology, the difference between the proportion of deductive confirmation terms and the proportion of inductive confirmation terms is statistically significant ($z = -8.16$, $p = 0.00$, two-sided). In Psychology, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 13.91$, $p = 0.00$, two-sided). In Sociology, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 7.64$, $p = 0.00$, two-sided). These results suggest that inductive confirmation terms are invoked significantly more often than deductive confirmation terms in Psychology and Sociology, but not

in Anthropology, publications that contain discussions of theories. In Anthropology publications that contain discussions of theories rather than hypotheses, deductive confirmation terms are invoked significantly more often than inductive confirmation terms.

In Computer Science, the difference between the proportion of deductive confirmation terms and the proportion of inductive confirmation terms is not statistically significant ($z = -1.12$, $p = 0.26$, two-sided). This result suggests that there is no significant difference between the frequency with which deductive and inductive confirmation terms are invoked in Computer Science publications that contain discussions of theories. In Mathematics, the difference between the proportion of deductive confirmation terms and the proportion of inductive confirmation terms is statistically significant ($z = -13.84$, $p = 0.00$, two-sided). This result suggests that deductive confirmation terms are invoked significantly more often than inductive confirmation terms in Mathematics publications that contain discussions of theories. In Statistics, the difference between the proportion of inductive confirmation terms and the proportion of deductive confirmation terms is statistically significant ($z = 27.402$, $p = 0.00$, two-sided). This result suggests that inductive confirmation terms are invoked significantly more often than deductive confirmation terms in Statistics publications that contain discussions of theories.

To check that the search methodology described in Section 2 returns genuine instances of the phenomenon in question (namely, instances of deductive and/or inductive confirmation terms in the context of talk about theories), I have selected three search results from the ‘theory’ corpus at random (emphasis added):

1. Life Sciences: “we tested the *prediction* from kin selection *theory*” (Smith et al. 2012, p. S440).

2. Physical Sciences: “they found an average rate for the 213 s complex in general agreement with the *expectation of cooling theory*” (Fontaine and Brassard 2008, p. 1079).
3. Social Sciences: “One-dimensional decision making is as likely a *consequence*, for example, of resilience *theory* and socio-biology as it is of environmental economics and environmental history” (Persson et al. 2018, p. 14).
4. Formal Sciences: “this *theory* carries with it its own *refutation*” (Rogers 1870, p. 248).

These instances of deductive and/or inductive confirmation terms in research articles published in scientific journals also provide context to the statistical results reported above. They illustrate how practicing scientists use deductive and/or inductive confirmation terms when they talk about theories in scholarly scientific practice.

4. Discussion

The results reported in Section 3 allow us to formulate tentative answers to the research questions of this empirical study of hypothesis testing in scientific practice. As far as (Q1) is concerned, the results of this empirical study show that practicing scientists use deductive terms when they talk about testing hypotheses and/or theories in scientific publications. As far as (Q2) is concerned, the results of this empirical study show that practicing scientists use inductive terms when they talk about testing hypotheses and/or theories in scientific publications. Based on the results of this empirical study, then, we can say that practicing scientists use both deductive and inductive confirmation terms when they talk about hypothesis testing in scientific publications. When we compare the proportions of deductive and inductive confirmation terms used within the scientific subjects tested in this study, however, we see that there are interesting and significant differences.

In the ‘hypothesis’ corpus, inductive confirmation terms are invoked significantly more often than deductive confirmation terms in life science publications. Since we observe the same pattern in the ‘theory’ corpus as well, we can say with some confidence that there is an emphasis on the inductive aspects of confirmation in the life sciences. In other words, the results from the ‘hypothesis’ corpus and the ‘theory’ corpus provide some empirical support for (H2) insofar as the life sciences are concerned. Clearly, if (H2) were true, it would explain the significantly higher proportions of inductive confirmation terms than deductive confirmation terms in life science publications. However, (H2) does not explain why there is an emphasis on the inductive aspects of confirmation in the life sciences to begin with. One might wonder, then, why there would be such an emphasis on the inductive aspects of confirmation in the life sciences in the first place. There may be several reasons for this. For example, some philosophers have argued that there are no nomological explanations in the life sciences. That is, unlike the physical sciences, which feature physical laws, such as the law of gravitation, the life sciences do not feature biological laws. For some, this is evidence that the life sciences are fundamentally different from the physical sciences. For others, this is evidence that identifying biological laws is just much more difficult than identifying physical laws. (See Rosenberg and McShea 2008, pp. 32-62.) If hypotheses and theories in the life sciences are fundamentally different from those in the physical sciences, then that could explain why there is an emphasis on the inductive aspects of confirmation in the life sciences but not in the physical sciences. Since the results of this empirical study do not bear on this question, however, it is better to leave this question to future studies.

As in the life sciences, inductive confirmation terms are invoked significantly more often than deductive confirmation terms in social science publications that contain discussions of

hypotheses. However, we cannot conclude from this that there is an emphasis on the inductive aspects of confirmation in the social sciences as in the life sciences. This is because we observe a different pattern in the ‘theory’ corpus. In the ‘theory’ corpus, inductive confirmation terms are invoked significantly more often than deductive confirmation terms in Psychology and Sociology, but not in Anthropology, publications that contain discussions of theories. In Anthropology publications that contain discussions of theories rather than hypotheses, deductive confirmation terms are invoked significantly more often than inductive confirmation terms. In other words, the results from the ‘hypothesis’ corpus and the ‘theory’ corpus provide some empirical support for (H2) but only insofar as Psychology and Sociology are concerned. These findings suggest two interesting possibilities that are worth pursuing in future studies: (a) that the social sciences are not a monolithic whole, methodologically speaking, and (b) that there may be a significant difference between hypothesis testing and theory confirmation in Anthropology. Indeed, the difference between the proportion of inductive confirmation terms in the ‘hypothesis’ corpus and deductive confirmation terms in the ‘theory’ corpus is statistically significant ($z = 4.81, p = 0.00$, two-sided).

While the results of this empirical survey suggest that there is an emphasis on the inductive elements of confirmation in the life sciences, as well as in Psychology and Sociology, but not in Anthropology, the same cannot be said about the physical sciences. In the ‘hypothesis’ corpus, there is no significant difference between the frequency with which deductive and inductive confirmation terms are invoked in physical science publications. In the ‘theory’ corpus, inductive confirmation terms are invoked significantly more often than deductive confirmation terms in Astronomy and Physics publications, but not in Chemistry publications. These mixed results, then, do not point to a clear pattern that would allow us to say with some confidence that

there is an emphasis on either the deductive or the inductive elements of confirmation in the physical sciences. In other words, the results from the ‘hypothesis’ corpus and the ‘theory’ corpus do not provide empirical support for either (H1) or (H2) insofar as the physical sciences are concerned.

Like the physical sciences, the formal sciences exhibit mixed patterns across subjects as well. In both the ‘hypothesis’ corpus and the ‘theory’ corpus, there is no significant difference between the frequency with which deductive and inductive confirmation terms are invoked in Computer Science publications. From these results, then, we cannot conclude that there is an emphasis on either the deductive or the inductive aspects of confirmation in Computer Science. By contrast, in both the ‘hypothesis’ corpus and the ‘theory’ corpus, the differences between the proportions of deductive confirmation terms and the proportions of inductive confirmation terms are statistically significant in Mathematics and in Statistics. Since we observe the same patterns in the ‘hypothesis’ corpus and the ‘theory’ corpus, we can say with some confidence that there is an emphasis on the deductive aspects of confirmation in Mathematics and an emphasis on the inductive aspects of confirmation in Statistics. In other words, the results from the ‘hypothesis’ corpus and the ‘theory’ corpus provide some empirical support for (H1) insofar as Mathematics is concerned but for (H2) insofar as Statistics is concerned. Clearly, if (H1) were true, it would explain the significantly higher proportions of deductive confirmation terms than inductive confirmation terms in Mathematics publications. However, (H1) does not explain why there is an emphasis on the deductive aspects of confirmation in Mathematics to begin with. One might wonder, then, why there would be such an emphasis on the deductive aspects of confirmation in Mathematics in the first place. There may be several reasons for this. For example, “if the paradigm of deductive reasoning is mathematical proof” (Glymour 1992, p. 6), then that could

explain why there is an emphasis on the deductive aspects of confirmation in Mathematics, but not why there is an emphasis on the inductive aspects of confirmation in Statistics. In that respect, these findings suggest two interesting possibilities that are worth pursuing in future studies: (a) that the formal sciences are not a monolithic whole, methodologically speaking, and (b) that there may be a significant difference between hypothesis testing and theory confirmation in Statistics versus hypothesis testing and theory confirmation in Mathematics.

All of the aforementioned differences between scientific subjects and across contexts (i.e., between hypothesis testing and theory confirmation) insofar as confirmation (or hypothesis testing) in scientific practice is concerned could be construed as providing some empirical evidence against the idea of methodological unity in science (see Cat 2017). That is, contrary to Popper's "central tenet of the unity of science [...] that testing of hypotheses [is] always to be conducted in the same manner as that of the natural scientist" (MacDonald 2004, p. 33), the results of this empirical study suggest that practicing scientists of various disciplines across the life, social, physical, and formal sciences think of hypothesis testing and theory confirmation in different terms (i.e., in terms of deductive or inductive confirmation terms). Admittedly, to the best of my knowledge, there are few proponents of Popper's thesis of the unity of scientific method in contemporary philosophy of science (see Verdugo 2009). Nevertheless, it would be interesting, I submit, to investigate the methodological differences between the formal, life, physical, and social sciences that the results of this empirical study suggest. In that respect, I submit, additional empirical studies are needed in order to understand confirmation (or hypothesis testing) in scientific practice, particularly, the differences between hypothesis testing in the life and social sciences versus the formal and physical sciences, and the differences within the sciences between hypothesis testing and theory confirmation.

As discussed in Section 2, like the results of other empirical studies, the results of this empirical study are not to be interpreted as conclusive evidence for or against any hypothesis about confirmation (or hypothesis testing) in scientific practice. Rather, they are supposed to contribute to our understanding of confirmation (or hypothesis testing) in science. Some philosophers of science who prefer rational reconstructions of science (see, e.g., Lakatos 1971, pp. 91-136),⁵ as opposed to empirical studies of scientific practices, might object that we do not gain much understanding of science by studying scientific practices, i.e., by studying what practicing scientists say and do.⁶ This is a methodological debate about how to do philosophy of science that is beyond the scope of this paper. For the purposes of this empirical study, I take it as a methodological assumption that we can gain valuable insights about science from what scientists say and do, specifically, what they say and do in their scholarly publications. For, as van Fraassen (1994, p. 184) puts it, “Any philosophical view of science is to be held accountable to actual scientific practice, scientific activity.” Accordingly, philosophical views of scientific confirmation (or hypothesis testing) should be held accountable to actual scientific practice. Assuming that what practicing scientists say and do in their research articles falls under “actual scientific practice” or “scientific activity,” it follows that philosophical views of scientific confirmation (or hypothesis testing) should be held accountable to what practicing scientists say and do in their research articles. The aim of this empirical study has been to shed light on what practicing scientists say and do in their research articles as far as confirmation (or hypothesis testing) is concerned.

⁵ According to Machery (2016, p. 480), “Rational reconstructions reconstruct the way scientists use particular concepts [or methods].”

⁶ Although, according to Lakatos (1971, p. 91), “any rational reconstruction of history needs to be supplemented by an empirical (socio-psychological) ‘external history’.”

5. Conclusion

The aim of this paper has been to contribute to our understanding of scientific confirmation (or hypothesis testing) in scientific practice by taking an empirical approach. I have used the tools of data science and corpus linguistics to study how practicing scientists talk about hypothesis testing or theory confirmation in research articles published in scientific journals. Overall, the results of this empirical survey suggest that there is an emphasis on mostly the inductive aspects of confirmation in the life sciences and the social sciences (with the exception of Anthropology), but not in the physical and the formal sciences. The results also point to interesting and significant differences between the scientific subjects within these disciplinary groups that are worth investigating in future studies. The significance of these findings is in providing empirical evidence against which to test our philosophical accounts of scientific confirmation (or hypothesis testing). For, as Machery (2016, p. 480) puts it, “if we can show experimentally that a candidate rational reconstruction [or philosophical view] of a given concept [or method] x has nothing or little to do with scientists’ unreconstructed use of x , then this gives us a strong reason to assume that the reconstruction is erroneous.”

Acknowledgments

I am very grateful to two anonymous reviewers of *International Studies in the Philosophy of Science* for their helpful comments on earlier drafts of this paper.

References

Andersen, H. and Hepburn, B. (2016). Scientific Method. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition).

<https://plato.stanford.edu/archives/sum2016/entries/scientific-method/>.

Barnes, E. C. (2018). Prediction versus Accommodation. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition).

<https://plato.stanford.edu/archives/fall2018/entries/prediction-accommodation/>.

Beebe, J. R. and Dellsén, F. (2020). Scientific Realism in the Wild: An Empirical Study of Seven Sciences and History and Philosophy of Science. *Philosophy of Science* 87 (2): 336-364.

Cat, J. (2017). The Unity of Science. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). <https://plato.stanford.edu/archives/fall2017/entries/scientific-unity/>.

Collard, M., Buchanan, B., and O'Brien, M. J. (2013), Alternative Pathways to Complexity: Evolutionary Trajectories in the Middle Paleolithic and Middle Stone Age. *Current Anthropology* 54 (S8): S388-S396.

Crupi, V. (2016). Confirmation. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). <https://plato.stanford.edu/archives/win2016/entries/confirmation/>.

De Marco, O. (2009). The Origin and Shaping of Planetary Nebulae: Putting the Binary Hypothesis to the Test. *Publications of the Astronomical Society of the Pacific* 121 (878): 316-342.

De Silva, N. (2010). A Concise, Elementary Proof of Arzela's Bounded Convergence Theorem. *The American Mathematical Monthly* 117 (10): 918-920.

Fontaine, G., and P. Brassard, P. (2008). The Pulsating White Dwarf Stars. *Publications of the Astronomical Society of the Pacific* 120 (872): 1043-1096.

Glymour, C. (1992). *Thinking Things Through: An Introduction to Philosophical Issues and Achievements*. Cambridge, MA: The MIT Press.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Fourth Edition. Cambridge, MA: Harvard University Press.

Howson, C. (1984). Bayesianism and Support by Novel Facts. *The British Journal for the Philosophy of Science* 35 (3): 245-251.

Hoyningen-Huene, P. (2008). Systematicity: The Nature of Science. *Philosophia* 36 (2): 167-180.

Kuipers, T. A. F. (2009). Empirical Progress and Truth Approximation by the ‘Hypothetico-Probabilistic Method’. *Erkenntnis* 70 (3): 313-330.

Lakatos, I. (1971). History of Science and Its Rational Reconstructions. In R. C. Buck and R. S. Cohen (eds.), *PSA 1970. Boston Studies in the Philosophy of Science, Vol. 8* (pp. pp 91-136). Dordrecht: Springer.

Lange, M. (2001). The Apparent Superiority of Prediction to Accommodation: a Reply to Maher. *The British Journal for the Philosophy of Science* 52 (3): 575-588.

London, B. (1992). School-Enrollment Rates and Trends, Gender, and Fertility: A Cross-National Analysis. *Sociology of Education* 65 (4): 306-316.

Machery, E. (2016). Experimental philosophy of science. In J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy* (pp. 475-490). New York: Wiley Blackwell.

MacDonald, G. (2004). The Grounds of Anti-Historicism. In A. O’Hear (ed.), *Karl Popper: Critical Assessments of Leading Philosophers Vol. IV: Politics and Social Science* (pp. 31-45). London: Routledge.

Maher, P. (1988). Prediction, Accommodation, and the Logic of Discovery. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1988 (1): 273-285.

Marcum, J. (2008). Instituting Science: Discovery or Construction of Scientific Knowledge? *International Studies in the Philosophy of Science* 22 (2): 185-210.

Mizrahi, M. (2013). The Pessimistic Induction: A Bad Argument Gone Too Far. *Synthese* 190 (15): 3209-3226.

Mizrahi, M. (2016). The History of Science as a Graveyard of Theories: A Philosophers' Myth? *International Studies in the Philosophy of Science* 30 (3): 263-278.

Mizrahi, M. (2020). The Case Study Method in Philosophy of Science: An Empirical Study. *Perspectives on Science* 28 (1): 63-88.

Persson, J., Hornborg, A., Olsson, L., and Thorén, H. (2018). Toward an alternative dialogue between the social and natural sciences. *Ecology and Society* 23 (4): 14.

Pitt, J. C. (2001). The Dilemma of Case Studies: Toward a Heraclitian Philosophy of Science. *Perspectives on Science* 9 (4): 373-382.

Potochnik, A., Colombo, M., and Wright, C. (2019). *Recipes for Science: An Introduction to Scientific Methods and Reasoning*. New York: Routledge.

Reiss, J. (2017). On the Causal Wars. In H. K. Chao and J. Reiss (eds.), *Philosophy of Science in Practice: Nancy Cartwright and the Nature of Scientific Reasoning* (pp. 45-67). Cham, Switzerland: Springer.

Rogers, J. E. T. (1870). On the Incidence of Local Taxation. *Journal of the Statistical Society of London* 33 (2): 243-263.

- Rosenberg, A. (2000). *Philosophy of Science: A Contemporary Introduction*. London: Routledge.
- Rosenberg, A. and McShea, D. W. (2008). *Philosophy of Biology: A Contemporary Introduction*. New York: Routledge.
- Ruan, L., and Yuan, M. (2011). An Empirical Bayes' Approach to Joint Analysis of Multiple Microarray Gene Expression Studies. *Biometrics* 67 (4): 1617-1626.
- Russell, H. N. (1919). On the Sources of Stellar Energy. *Publications of the Astronomical Society of the Pacific* 31 (182): 205-211.
- Salmon, W. C. (1970). Bayes's Theorem and the History of Science. In R. H. Stuewer (ed.), *Historical and Philosophical Perspectives of Science* (pp. 68-86). New York: Gordon and Breach.
- Salmon, M. H. (1976). "Deductive" versus "Inductive" Archeology. *American Antiquity* 41 (3): 376-381.
- Salmon, M. H. (1982). *Philosophy and Archeology*. New York: Academic Press.
- Smith, J. E., Swanson, E. M., Reed, D., and Holekamp, K. E. (2012). Evolution of Cooperation among Mammalian Carnivores and Its Relevance to Hominin Evolution. *Current Anthropology* 53 (S6): S436-S452.
- Society for Philosophy of Science in Practice. (2006-2019). Mission Statement. *Society for Philosophy of Science in Practice*. Accessed November 1, 2019. <https://philosophy-science-practice.org/about/mission-statement>.

Soler, L., Zwart, S., Lynch, M., and Israel-Jost, V. (Eds.). (2014). *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*. London: Routledge.

Tedbury, P. R., Novikova, M., Ablan, S. D., and Freed, E. O. (2016). Biochemical evidence of a role for matrix trimerization in HIV-1 envelope glycoprotein incorporation. *Proceedings of the National Academy of Sciences of the United States of America* 113 (2): E182-E190.

Treves, A. and Chapman, C. A. (1996). Conspecific Threat, Predation Avoidance, and Resource Defense: Implications for Grouping in Langurs. *Behavioral Ecology and Sociobiology* 39 (1): 43-53.

Trijsburg, R. W., Bal, J. A., Parsowa, W. P., Erdman, R. A. M., and Duivenvoorden, H. J. (1989). Prediction of Physical Indisposition with the Help of a Questionnaire for Measuring Denial and Overcompensation. *Psychotherapy and Psychosomatics* 51 (4): 193-202.

Van Fraassen, B. C. (1994). Gideon Rosen on Constructive Empiricism. *Philosophical Studies* 74 (2): 179-192.

Verdugo, C. (2009). Popper's Thesis of the Unity of Scientific Method: Method versus Techniques. In Z. Parusniková and R. S. Cohen (eds.), *Rethinking Popper* (pp. 155-160). Dordrecht: Springer.

Wills, A. J. (2009). Prediction Errors and Attention in the Presence and Absence of Feedback. *Current Directions in Psychological Science* 18 (2): 95-100.

Zorn, D. M. (2004). Here a Chief, There a Chief: The Rise of the CFO in the American Firm. *American Sociological Review* 69 (3): 345-364.