

Philosophical Reasoning about Science: A Quantitative, Digital Study

Moti Mizrahi, Florida Institute of Technology
Mike Dickinson, University of Illinois

Forthcoming in *Synthese*

Abstract: In this paper, we set out to investigate the following question: if science relies heavily on induction, does philosophy of science rely heavily on induction as well? Using data mining and text analysis methods, we study a large corpus of philosophical texts mined from the JSTOR database (n = 14,199) in order to answer this question empirically. If philosophy of science relies heavily on induction, just as science supposedly does, then we would expect to find significantly more inductive arguments than deductive arguments and abductive arguments in the published works of philosophers of science. Using indicator words to classify arguments by type (namely, deductive, inductive, and abductive arguments), we search through our corpus to find patterns of argumentation. Overall, the results of our study suggest that philosophers of science do rely on inductive inference. But induction may not be as foundational to philosophy of science as it is thought to be for science, given that philosophers of science make significantly more deductive arguments than inductive arguments. Interestingly, our results also suggest that philosophers of science do not rely on abductive arguments all that much, even though philosophers of science consider abduction to be a cornerstone of scientific methodology.

Keywords: abduction; argument; corpus analysis; deduction; digital studies of science; induction; metaphilosophy of science; text mining

1. Introduction

In this paper, we set out to investigate the following question empirically: if science relies heavily on induction, as philosophers of science believe, does philosophy of science rely heavily on induction as well? As Okasha (2016, p. 19) observes, “Most philosophers think it’s obvious that science relies heavily on induction, indeed so obvious that it hardly needs arguing for.” Likewise, Henderson (2020) points out that “it is also generally thought that [inductive inference] is at the very foundation of the scientific method.” This is why, for some philosophers of science, “an adequate defence of induction was central to the defence of the rationality of reasoning in science” (Gower 1997, p. 189). As Russell (1912, p. 107) puts it:

The general principles of science, such as the belief in the reign of law, and the belief that every event must have a cause, are as completely dependent upon *the inductive principle* as are the beliefs of daily life. All such general principles are believed because mankind have found innumerable instances of their truth and no instances of their falsehood. But this affords no evidence for their truth in the future, unless *the inductive principle* is assumed (emphasis added).

In other words, if inductive reasoning is a central part of scientific reasoning, then the rationality of scientific reasoning depends in part on a rational justification for induction. Finally, according to Douglas (2017, p. 86):

The most important thing to understand about science is its jointly critical and *inductive* nature. [...] Because the theories always say more than the available evidence, the evidence provides at best *inductive* and thus incomplete support for the theories. [...] explanations and theories never have complete empirical support, yet the primary mode of support is empirical. It is in this sense that *science is inductive*” (emphasis added).¹

All of this leads us to the aforementioned question: if “science relies heavily on induction” (Okasha 2016, p. 19), does philosophy of science rely heavily on induction as well? That is, if inductive inference “is at the very foundation of the scientific method” (Henderson 2020), is it also at the very foundation of philosophical reasoning about science? In other words, do those who study science (namely, philosophers of science) rely on inductive arguments just as those whom they study (namely, scientists) do? This is the research question that guides our empirical study. Using data mining and text analysis methods, we study a large corpus of philosophical texts mined from the JSTOR database (n = 14,199) in order to answer this question empirically. If philosophy of science relies heavily on induction, just as science supposedly does, or if philosophy of science is inductive, just as science supposedly is, then we would expect to find significantly more inductive arguments than deductive arguments and abductive arguments in the published works of philosophers of science.

Using indicator words to classify arguments by type (namely, deductive, inductive, and abductive arguments), we search through our digital corpus to find patterns of argumentation. Specifically, we search for deductive, inductive, and abductive arguments in articles published in the following philosophy of science journals: *British Journal for the Philosophy of Science* (BJPS), *History and Philosophy of the Life Sciences* (HPLS), *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *Journal for General Philosophy of Science* (JGPS), *Philosophy of Science*, and the *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (PSA). We conducted these searches allowing for three, six, and ten words between indicator words for arguments (e.g., ‘therefore’, ‘hence’, and the like) and indicator words for argument types (e.g., ‘necessarily’ for deductive arguments, ‘probably’ for inductive arguments, and ‘best explain’ for abductive arguments) in order to find out how prevalent deductive, inductive, and abductive arguments are in articles published in philosophy of science journals.

Before we report the results of our empirical study in Section 3, we describe our methodology in more detail in Section 2. (See also Appendix I for text mining methods in R and Appendix II for a check of interrater reliability.) In Section 4, we discuss the results of our quantitative, digital study. Overall, the results of our study suggest that philosophers of science do rely on inductive inference in their published works. But induction may not be as foundational to philosophy of science as it is thought to be for science, given that philosophers of science

¹ See also (Brigandt 2014, p. 254): “Assuming that confirmation in science is inductive, such a logic of induction describes the *form* of the confirmation relation between evidence statements and theory, abstracting away from the particular empirical *content* involved in a particular instance of confirmation” (emphasis in original). In that respect, using the digital methods of text mining and corpus analysis, similar to the methods we have used in this empirical study, Mizrahi (2020) finds 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 sciences and the formal sciences.

make significantly more deductive arguments than inductive arguments in articles published in philosophy of science journals. Interestingly, the results of our study also suggest that philosophers of science do not rely on abductive arguments all that much, even though philosophers of science consider abduction to be a cornerstone of scientific methodology.

2. Methods

Introductory textbooks to logic and argumentation typically contain a brief discussion of indicator words. Indicator words are “words or phrases that typically serve to signal the appearance of an argument’s conclusion or of its premises” (Copi et al. 2014, 11). More specifically, there are premise indicators, which include words like ‘because’ and phrases like ‘inferred from’ and the like. Premise indicators indicate a premise of an argument. In addition, there are conclusion indicators, which include words like ‘therefore’ and phrases like ‘it follows that’ and the like. Conclusion indicators indicate a conclusion of an argument. For example, Morrow and Weston (2011, p. 5) instruct students to look for indicator words in order to distinguish between premises and conclusions as follows:

Some words or phrases are *conclusion indicators*. These are words or phrases that tell you that you’re about to read or hear the conclusion of an argument. Other words or phrases are *premise indicators*. These tell you that you’re about to read or hear a premise (emphasis in original).

Morrow and Weston (2011, p. 5) go on to provide a list of premise indicators, which includes words like ‘because’ and ‘this follows from’, and a list of conclusion indicators, which includes words like ‘therefore’ and ‘hence’. Similarly, in her introductory book on logic and argumentation, Govier (2013, p. 4) writes, “Indicator words suggest the presence of argument and help to indicate its structure. Some indicator words, like *therefore*, come before the conclusion in an argument. Other indicator words, like *since* and *because*, come before premises.” Govier’s (2013, pp. 4-5) list of premise indicators includes the following words and phrases: ‘since’, ‘because’, ‘for’, ‘as indicated by’, ‘follows from’, ‘may be inferred from’, ‘may be derived from’, ‘on the grounds that’, ‘for the reason that’, ‘as shown by’, ‘given that’, and ‘may be deduced from’. And her list of conclusion indicators includes the following words and phrases: ‘therefore’, ‘thus’, ‘so’, ‘consequently’, ‘hence’, ‘then’, ‘it follows that’, ‘it can be inferred that’, ‘in conclusion’, ‘accordingly’, ‘for this reason (or for all these reasons) we can see that’, ‘on these grounds it is clear that’, ‘proves that’, ‘shows that’, ‘indicates that’, ‘we can conclude that’, ‘we can infer that’, and ‘demonstrates that’ (Govier 2013, pp. 5-6).

In addition to helping students identify the premises and conclusions of arguments, indicators also help students distinguish between different types of arguments, such as deductive arguments and inductive arguments. For example, according to Baronett (2016, p. 23):

To help identify arguments as either deductive or inductive, one thing we can do is look for key words or phrases. For example, the words “necessarily,” “certainly,” “definitely,” and “absolutely” suggest a deductive argument. [...] [This is because a] *deductive argument* is one in which it is claimed that the conclusion follows *necessarily* from the premises. [...] On the other hand, the words “probably,” “likely,” “unlikely,”

“improbable,” “plausible,” and “implausible” suggest inductive arguments. [...] [This is because an] *inductive argument* is one in which it is claimed that the premises make the conclusion *probable* (emphasis in original).

Similarly, according to Hurley and Watson (2018, p. 34), “In deciding whether an argument is inductive or deductive, we look to certain objective features of the argument” (2018, pp. 34-35). One of those objective features is “the occurrence of special indicator words” (Hurley and Watson 2018, pp. 34-35). According to Hurley and Watson (2018, p. 35), “inductive indicators” include words and phrases such as ‘probably’, ‘improbable’, ‘plausible’, ‘implausible’, ‘likely’, ‘unlikely’, and ‘reasonable to conclude’, whereas “deductive indicators” include words and phrases such as ‘it necessarily follows that’, ‘certainly’, ‘absolutely’, and ‘definitely’.²

Now, we can use these deductive indicators and inductive indicators to look for deductive arguments and inductive arguments in philosophical texts in much the same way that students use them to look for arguments in any text. In that respect, we are following Ashton and Mizrahi’s (2018) methodology, but with a novel addition. That is, to the aforementioned deductive and inductive indicator words, we have added indicator words for abductive arguments, i.e., arguments in which the conclusion is supposed to be the best explanation for some phenomenon (Govier 2013, pp. 298-302). Broadly speaking, abductive arguments may be considered inductive arguments insofar as the premises of an abductive argument are intended to make its conclusion *probably*, but not *necessarily*, true.³ So, if “[a]n *inductive argument* is one in which it is claimed that the premises make the conclusion *probable*” (Baronett 2016, p. 23; emphasis in original), and the premises of abductive arguments are intended to provide *probable* support for their conclusions, then abductive arguments can be considered a type of inductive arguments. Nevertheless, some philosophers and logicians treat abductive arguments as a distinct type of argumentation. Indeed, Baronett (2016) himself discusses abduction and Inference to the Best Explanation (IBE) in a chapter titled “Causality and Scientific Arguments,” which is separate from the chapters on deduction and induction in his logic textbook. According to Baronett (2016, p. 652), “In *inference to the best explanation*, we reason from the premise that a hypothesis would explain certain facts to the conclusion that the hypothesis is the best explanation for those facts” (emphasis in original). Accordingly, abductive indicators include words and phrases such as ‘account for’, ‘best explain’, ‘make sense of’, and ‘best explanation for’ (Overton 2013). The types of arguments we searched for in this quantitative, digital study and their associated indicators are listed in Table 1.

Table 1. Types of arguments and their indicator words with examples from philosophical texts

² According to Salmon (2013), “Expressions such as *must*, *it must be the case that*, *necessarily*, *inevitably*, *certainly*, and *it can be deduced that* frequently indicate that an argument is deductive,” (p. 86), whereas expressions such as “*probably*, *usually*, *tends to support*, *likely*, *very likely*, and *almost always*” typically indicate that an argument is inductive (p. 94).

³ Cf. Ashton and Mizrahi (2018, p. 61), footnote 4.

Argument Types	Indicators	Examples
<i>Abductive</i>	account for, best explain, make sense of, best explanation for	“We infer that middle-sized objects exist, because their existence <i>provides the best explanation for</i> the patterns in our sense experience” (Trout 1998, p. 97).
<i>Deductive</i>	absolutely, certainly, definitely, necessarily	“if, as he says, such an infinite series really is impossible then it does <i>absolutely follow that</i> if anything exists in time at all, there must have been a moment, before which nothing existed” (Moore 1954, p. 175).
<i>Inductive</i>	likely, unlikely, probably, improbable	“There are instances of intense suffering that are inscrutable [...]; <i>so, probably</i> those instances of suffering are gratuitous; <i>therefore, probably</i> , God does not exist” (Wielenberg 2015, p. 304).

Of course, as logic textbooks will typically mention as well, we have to keep in mind that these abductive, deductive, and inductive indicators are no more than *indicators*. That is, they are not sure signs for the presence (or absence) of arguments in texts. As Hurley (2016, p. 23) puts it, “the mere occurrence [or absence] of an indicator word does not *guarantee* the presence [or absence] of an argument” (emphasis added). Nevertheless, indicator words are still useful and reliable indicators for the presence of arguments in text, which is why students of logic and argumentation are instructed to look for them. As Lepore and Cumming (2013, p. 6) put it, “Although there are no sure signs of whether an argument is present, fairly reliable indicators exist.” Lepore and Cumming (2013, p. 6) proceed to list some of the aforementioned indicator words as well (see Table 1). In addition, since our aim is to study arguments made by academic philosophers, which are published in academic journals of philosophy, specifically, academic journals of philosophy of science, and academic “philosophers are careful folk, trained in the ways of argument” (Currie 2016, p. 200), we can be quite confident that, as professional arguers, academic philosophers rarely misuse indicators in an effort to make non-arguments appear as arguments (see also Ashton and Mizrahi 2018, p. 62).

The quantitative methods we use in this digital study, namely, data mining and text analysis, allow us to overcome the limitations of relying on selective quotation. After all, one can easily find instances of the aforementioned indicator words in philosophical texts (see Table 1). However, selected quotations may or may not be representative of academic philosophy of science as a whole. By using data mining and text analysis methods, we can study a large corpus of philosophy of science texts, and thus obtain a broader view of the argumentative landscape in academic philosophy of science. Of course, empirical methodologies have limitations of their own. As far as the methods of data mining and text analysis are concerned, there are two major limitations. First, we can only study and analyze what is explicitly mentioned in the corpus. For the purpose of this quantitative, digital study, then, our corpus of philosophy of science texts must contain explicit mentions of the indicator words listed in Table 1, so that we can analyze ratios, means, and patterns of argumentation. It is reasonable to suppose that there would be such

explicit mentions of the indicator words listed in Table 1 in philosophy of science texts, given that academic philosophers of science are professional arguers; that is, “trained in the ways of argument” (Currie 2016, p. 200).

Second, as with empirical methodologies in general, there may be a few false positives and/or false negatives when it comes to our empirical methodology in particular. More explicitly, as far as the methods of text mining and analysis are concerned, false negatives could occur when we search for a specific word w in a corpus, but we do not find it because the corpus contains a synonym of w rather than w . For example, although unlikely, it is possible that our corpus of philosophy of science texts contains no instances of ‘probably’, and so a search for ‘probably’ would return zero search results, because academic philosophers of science use ‘likely’ instead of ‘probably’ in all the philosophy of science texts that make up our corpus. On the other hand, false positives could occur when we find instances of a word w in our corpus, but those instances contain irrelevant uses of w . For the purpose of this quantitative, digital study, then, the corpus of philosophy of science texts must contain not only explicit mentions of the abductive, deductive, and inductive indicators listed in Table 1, but also explicit mentions of those indicators in the context of argumentation. For example, instances of ‘certainly’ that occur outside of any argumentative context would be considered false positives for the purposes of this study.

Now, there are a few things that we can do in order to overcome the limitations of our digital, corpus-based approach. First, we can refine our searches by expanding our search terms to include as many indicator words as we can. In this study, we have four indicator words for each argument type (see Table 1). This search methodology is designed to minimize the number of false negatives, i.e., occurrences of abductive, deductive, and inductive arguments in philosophy of science texts that are indicated by words other than the common indicator words, such as ‘best explain’, ‘necessarily’, and ‘probably’, by using synonymous indicator words and phrases, such as ‘account for’, ‘certainly’, and ‘likely’.⁴

Second, we can further refine our searches by pairing the argument type indicators with indicator words for conclusions, such as ‘therefore’ and ‘hence’. Since our aim is to find out whether philosophy of science relies on induction, we need to find out what types of arguments academic philosophers of science actually make in philosophy of science publications. To this end, we need to search for the abductive, deductive, and inductive indicators listed in Table 1 in argumentative contexts by pairing the abductive, deductive, and inductive indicators listed in Table 1 with conclusion indicators, such as ‘therefore’ and ‘hence’. By anchoring the abductive, deductive, and inductive indicators listed in Table 1 to conclusion indicators, such as ‘therefore’ and ‘hence’, we can be quite confident that our indicators for argument types (see Table 1) actually indicate arguments in the corpus, given that an argument must have a conclusion, and thus that the number of false positives will be minimized. This procedure results in the argument indicator pairs listed in Table 2.

Table 2. Indicator pairs for deductive, inductive, and abductive arguments

⁴ In Hylnad’s (2005) taxonomy of metadiscourse signals, ‘probably’ and ‘likely’ are classified as hedges, whereas in Salager-Meyer’s (1994) taxonomy they are classified as shields. For a critical assessment of these taxonomies, see Thabet (2018).

Deductive indicator pairs	Inductive indicator pairs	Abductive indicator pairs
therefore necessarily	therefore probably	therefore account for
therefore certainly	therefore likely	therefore best explain
therefore definitely	therefore unlikely	therefore make sense of
therefore absolutely	therefore improbable	therefore best explanation for
hence necessarily	hence probably	hence account for
hence certainly	hence likely	hence best explain
hence definitely	hence unlikely	hence make sense of
hence absolutely	hence improbable	hence best explanation for
so necessarily	so probably	so account for
so certainly	so likely	so best explain
so definitely	so unlikely	so make sense of
so absolutely	so improbable	so best explanation for
consequently necessarily	consequently probably	consequently account for
consequently certainly	consequently likely	consequently best explain
consequently definitely	consequently unlikely	consequently make sense of
consequently absolutely	consequently improbable	consequently best explanation for
proves necessarily	proves probably	proves account for
proves certainly	proves likely	proves best explain
proves definitely	proves unlikely	proves make sense of
proves absolutely	proves improbable	proves best explanation for
thus necessarily	thus probably	thus account for
thus certainly	thus likely	thus best explain
thus definitely	thus unlikely	thus make sense of
thus absolutely	thus improbable	thus best explanation for
follows necessarily	follows probably	follows account for
follows certainly	follows likely	follows best explain
follows definitely	follows unlikely	follows make sense of
follows absolutely	follows improbable	follows best explanation for
accordingly necessarily	accordingly probably	accordingly account for
accordingly certainly	accordingly likely	accordingly best explain
accordingly definitely	accordingly unlikely	accordingly make sense of
accordingly absolutely	accordingly improbable	accordingly best explanation for
infer necessarily	infer probably	infer account for
infer certainly	infer likely	infer best explain
infer definitely	infer unlikely	infer make sense of
infer absolutely	infer improbable	infer best explanation for
conclude necessarily	conclude probably	conclude account for
conclude certainly	conclude likely	conclude best explain
conclude definitely	conclude unlikely	conclude make sense of
conclude absolutely	conclude improbable	conclude best explanation for

Third, we can stem some of the indicator words listed in Table 2 above, so as to minimize the number of false negatives as much as possible. For example, if we search our corpus for the

inductive indicator pair “therefore probably” (as in “therefore, probably p ”), we might miss inductive arguments where the conclusion is phrased along the lines of “therefore, it is probable that p .” In order to avoid that, we can stem the word ‘probably’, and thus make sure that our search results will include instances of “therefore probably” and “therefore probable” (with up to three, six, or ten words between ‘therefore’ and ‘probabl*’). Likewise, if we search our corpus for the abductive indicator pair “so best explain” (as in “so, it’s best to posit p to explain q ”), we might miss abductive arguments where the conclusion is phrased along the lines of “so, p best explains q ” or “so, q is best explained by p .” In order to avoid that, we can stem the word ‘explain’, and thus make sure that our search results will include instances of “so best explain,” “so best explains,” and “so best explained” (with up to three, six, or ten words between ‘so’ and ‘explain*’). We say more about how we stemmed some of the indicator words from Table 2 in Appendix I.

By searching for the indicator pairs listed in Table 2 in our corpus, with stemming (see Appendix I), we can find out what types of arguments academic philosophers of science make in their published works and with what frequency. For each of the pairs listed in Table 2, we ran three kinds of searches: (a) a search allowing for up to three words between argument type indicator, e.g., ‘necessarily’, and argument indicator, e.g., ‘therefore’, (b) a search allowing for up to six words between argument type indicator, e.g., ‘probably’, and argument indicator, e.g., ‘hence’, and (c) a search allowing for up to ten words between argument type indicator, e.g., ‘account for’, and argument indicator, e.g., ‘so’. This methodology is designed to help us find answers to our research question while minimizing the number of false positives and false negatives.

It is important to emphasize again that our search methodology is not totally immune from counting false negatives and/or false positives, as we discussed above. One reason to think that there might be some false negative results in our datasets is that academic philosophers of science could be omitting indicator words from their academic publications deliberately because they are writing for a professional audience of academic philosophers of science. Presumably, being academic philosophers of science themselves, readers of philosophy of science journals do not need indicator words to identify arguments in text. This is possible, of course, although omitting indicator words might seem to run counter to academic philosophers’ professed commitment to rigor and clarity in philosophical writing. For omitting indicators words would make it less clear to any reader, academic philosopher or not, where the argument in the text is, what type of argument is being made, and what the premises and the conclusion of the argument are. And academic philosophers, particularly those working in the analytic tradition, “pride themselves on skill in argumentation” (Rorty 2006, p. 70). As Lackey (2005, p. 277) puts it, “Analytic philosophers pride themselves on being logical, rigorous, and clear.”

To address our research question empirically using the corpus-based methods of this study, we need to be able to distinguish between not only types of arguments (namely, deductive, inductive, or abductive arguments) but also types of journals in our corpus. More specifically, we need to focus on philosophy of science journals that publish work in philosophy of science. Our corpus of philosophy texts contains text from articles published in the following philosophy of science journals:

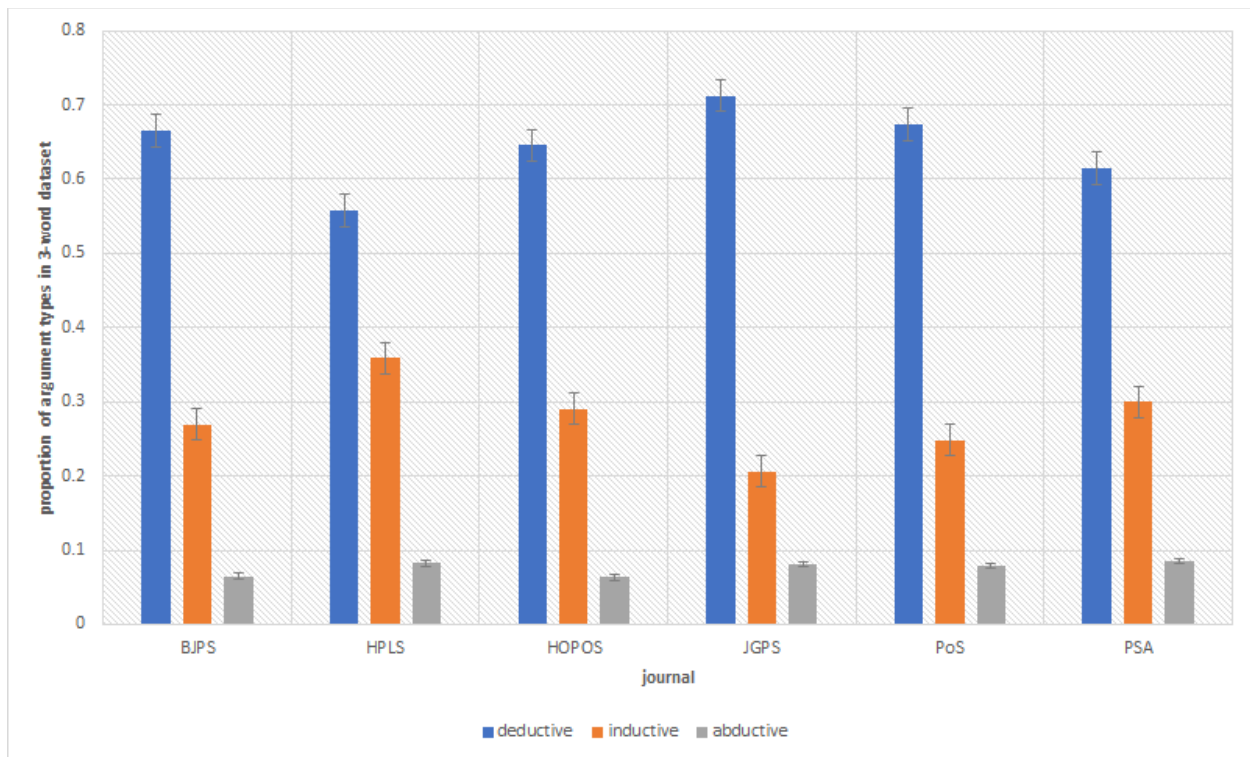
- *British Journal for the Philosophy of Science* (BJPS)
- *History and Philosophy of the Life Sciences* (HPLS)
- *HOPOS: The Journal of the International Society for the History of Philosophy of Science* (HOPOS)
- *Journal for General Philosophy of Science* (JGPS)
- *Philosophy of Science* (PoS)
- *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (PSA).

Our datasets contain articles published in the aforementioned philosophy of science journals between the years 1934 and 2014 (n = 14,199). By searching for the argument indicator pairs listed in Table 2 in articles published in the aforementioned philosophy of science journals, we can find out what types of arguments academic philosophers of science make in their published articles and with what frequency. This, in turn, will bring us a little closer to answering our research question: Does philosophy of science rely heavily on inductive inference? For more details on our text mining methods in R, see Appendix I.

3. Results

In searches permitting three words between argument indicator root and anchor, the proportions of deductive arguments are always higher than the proportions of inductive arguments and those of abductive arguments across all the philosophy of science journals comprising the corpus for this study (see Figure 1).

Figure 1. Proportions of argument types by philosophy of science journal (three-word dataset)



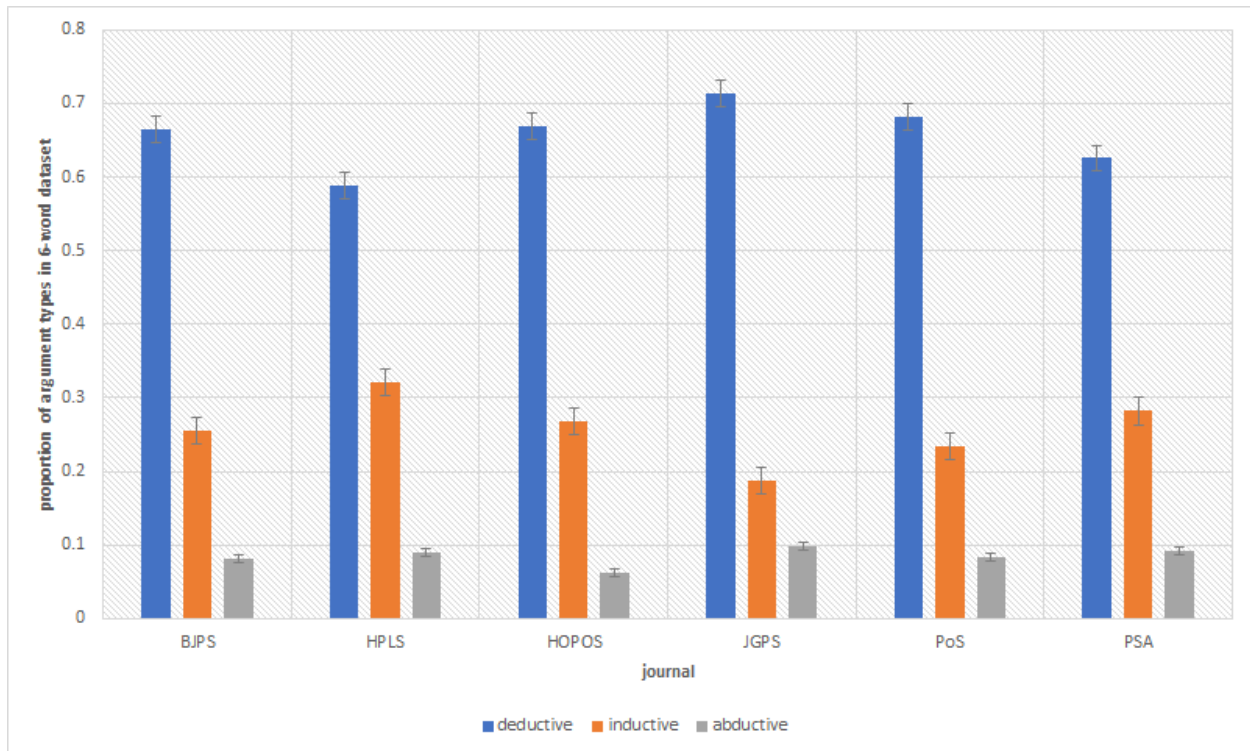
The results of a one-way ANOVA indicated that the proportions of deductive, inductive, and abductive arguments in the three-word dataset are unequal, $F(2, 15) = 266.41, p < .001$. So we conducted further two-sample z -tests in order to determine whether there are significant differences between the proportions of deductive arguments and inductive arguments in the three-word dataset. Across all the philosophy of science journals included in this study, the proportion of deductive arguments is significantly larger than the proportion of inductive arguments in the three-word dataset. For example, as far as articles published in the BJPS are concerned, the difference between the proportion of deductive arguments and the proportion of inductive arguments is statistically significant ($z = 31.49, p < .001$, two-sided). The results of these z -tests are summarized in Table 3.

Table 3. Results of z -tests comparing the proportions of deductive and inductive arguments by philosophy of science journal in the three-word dataset

Journal	Deductive proportion	Inductive proportion	z	p
BJPS	0.66	0.26	31.49	< .001
HPLS	0.55	0.35	6.72	< .001
HOPOS	0.64	0.29	5.26	< .001
JGPS	0.71	0.21	16.74	< .001
PoS	0.67	0.24	42.38	< .001
PSA	0.61	0.31	14.16	< .001

In searches permitting six words between argument indicator root and anchor, we find patterns that are similar to those found in our three-word dataset. That is, the proportions of deductive arguments are always higher than the proportions of inductive arguments and those of abductive arguments across the philosophy of science journals comprising the corpus for this study (see Figure 2).

Figure 2. Proportions of argument types by philosophy of science journal (six-word dataset)



The results of a one-way ANOVA indicated that the proportions of deductive, inductive, and abductive arguments in the six-word dataset are unequal, $F(2, 15) = 375.75$, $p < .001$. So we conducted further two-sample z -tests in order to determine whether there are significant differences between the proportions of deductive, inductive, and abductive arguments in the six-word dataset. Across all the philosophy of science journals included in this study, the proportion of deductive arguments is significantly larger than the proportion of inductive arguments in the six-word dataset, which is similar to the pattern exhibited by the data from our three-word dataset. For example, as far as articles published in *Philosophy of Science* are concerned, the difference between the proportion of deductive arguments and the proportion of inductive arguments is statistically significant ($z = 56.23$, $p < .001$, two-sided). The results of these z -tests are summarized in Table 4. These results are consistent with the results from the three-word dataset.

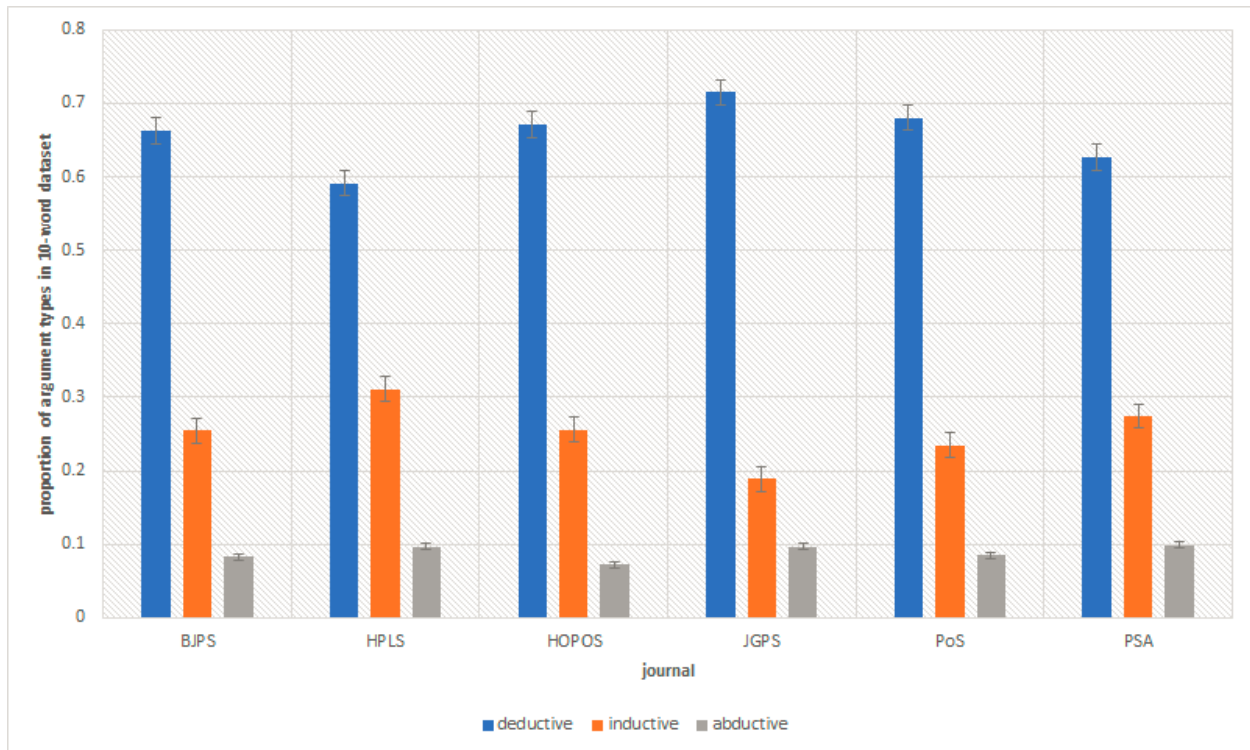
Table 4. Results of z -tests comparing the proportions of deductive and inductive arguments by philosophy of science journal in the six-word dataset

Journal	Deductive proportion	Inductive proportion	z	p
BJPS	0.66	0.25	40.85	< .001
HPLS	0.58	0.32	11.83	< .001
HOPOS	0.66	0.26	7.49	< .001

JGPS	0.71	0.18	21.81	< .001
PoS	0.68	0.23	56.23	< .001
PSA	0.62	0.28	19.51	< .001

In searches permitting ten words between argument indicator root and anchor, we find patterns that are similar to those found in our three-word and six-word datasets. That is, the proportions of deductive arguments are always higher than the proportions of inductive arguments and those of abductive arguments across the philosophy of science journals comprising the corpus for this study (see Figure 3).

Figure 3. Proportions of argument types by philosophy of science journal (ten-word dataset)



The results of a one-way ANOVA indicated that the proportions of deductive, inductive, and abductive arguments in the 10-word dataset are unequal, $F(2, 15) = 423.79, p < .001$. So we conducted further two-sample z -tests in order to determine whether there are significant differences between the proportions of deductive, inductive, and abductive arguments in the ten-word dataset. Across all the philosophy of science journals included in this study, the proportion of deductive arguments is significantly larger than the proportion of inductive arguments in the ten-word dataset, which is similar to the pattern exhibited by the data from our three-word and six-word datasets. For example, as far as articles published in the *Journal for General Philosophy of Science* are concerned, the difference between the proportion of deductive arguments and the proportion of inductive arguments is statistically significant ($z = 25.54, p <$

.001, two-sided). The results of these z -tests are summarized in Table 5. These results are consistent with the results from the three-word and six-word datasets.

Table 5. Results of z -tests comparing the proportions of deductive and inductive arguments by philosophy of science journal in the ten-word dataset

Journal	Deductive proportion	Inductive proportion	z	p
BJPS	0.66	0.25	47.91	< .001
HPLS	0.59	0.31	14.81	< .001
HOPOS	0.67	0.25	8.99	< .001
JGPS	0.71	0.18	25.54	< .001
PoS	0.68	0.23	65.83	< .001
PSA	0.62	0.27	23.61	< .001

4. Discussion

As we discussed in Section 1, our digital, corpus-based study was designed to address the following question empirically: if “science relies heavily on induction” (Okasha 2016, p. 19), does philosophy of science rely heavily on induction as well? That is, if inductive inference “is at the very foundation of the scientific method” (Henderson 2020), is it also at the very foundation of philosophical reasoning about science? If philosophy of science relies heavily on induction, just as science supposedly does (Okasha 2016, p. 19), or if philosophy of science is inductive, just as science supposedly is (Douglas 2017, p. 86), then we would expect to find significantly more inductive arguments than deductive arguments and abductive arguments in the published works of philosophers of science.

Overall, the results of our quantitative, digital study suggest that philosophy of science does rely on induction to some extent. For philosophers of science do make inductive arguments in their published articles. But induction may not be as foundational to philosophy of science as it is thought to be for science. For, in addition to inductive arguments, philosophers of science also make deductive arguments in their published articles. In fact, our results suggest that articles published in philosophy of science journals contain significantly more deductive arguments than inductive arguments and abductive arguments. These results, then, do not provide empirical support to the hypothesis that philosophy of science relies heavily on induction, just as science supposedly does. For, if philosophy of science were inductive, just as science supposedly is, then we would expect to find significantly more inductive arguments than deductive arguments and abductive arguments in articles published in philosophers of science journals. But that is not what we have found.

Interestingly, the results of our digital study suggest that philosophers of science do not rely on abductive inferences all that much. This is a surprising finding, we submit, because, in addition to thinking that “science relies heavily on induction” (Okasha 2016, p. 19), philosophers of science also tend to think that abductive inference or Inference to the Best Explanation (IBE) is “ubiquitous in scientific practice” (Chakravartty 2017). As McCain and Poston (2017, p. 1) put it:

Explanatory reasoning is quite common. Not only are *rigorous inferences to the best explanation (IBE) used pervasively in the sciences*, explanatory reasoning is virtually ubiquitous in everyday life. It is not a stretch to say that we implement explanatory reasoning in a way that is “so routine and automatic that it easily goes unnoticed” (Douven, 2011). (emphasis added)

As Douven (2017) observes in the *Stanford Encyclopedia of Philosophy* entry cited in the quote from McCain and Poston (2017, p. 1) above, “philosophers of science have argued that abduction is a cornerstone of scientific methodology.” McMullin (1992) calls abductive inference “the inference that makes science.” Abductive inference may be the inference that makes science, as McMullin (1992) says, but it does not seem to be the inference that makes philosophy of science. Our results suggest that the inference that makes philosophy of science is deduction, not induction or abduction.⁵

As an anonymous reviewer suggested, it would be useful to compare the argumentation patterns we have uncovered in philosophy of science publications to argumentation patterns in articles published in science journals. Comparing argumentation patterns in philosophy of science to those in science would help us find out whether, and to what extent, philosophers of science are emulating what they consider to be best argumentation practices in science. It would also allow us to test empirically what philosophers of science generally take for granted, namely, that “science is inductive” (Douglas 2017, p. 86). As Okasha (2016, p. 19) observes, “Most philosophers think it’s obvious that science relies heavily on induction, indeed so obvious that it hardly needs arguing for.” Unfortunately, our corpus consists of philosophy of science journal articles only. For this study, we do not have a corpus of science journal articles, so we cannot undertake the proposed comparison in this study. We have to leave that to future studies.

Another question for future studies, which was also suggested by an anonymous reviewer, is whether some arguments carry more weight than others. This question arises from the observation that there can be arguments within arguments (also known as “nested arguments”). For example, the premises of a deductive argument can themselves be supported by premises that, together with the premises they are meant to support, make up inductive arguments. In this case, we have two inductive arguments nested within a deductive argument. Then the question is how we should count the arguments. Do we have one deductive argument

⁵ In this study, we have not looked at changes over time. However, there is some empirical evidence to suggest that academic philosophy (and so, presumably, academic philosophy of science as well) is undergoing a methodological change of some sort. For example, using digital, corpus-based methods similar to the ones used in the present study, Ashton and Mizrahi (2018) find evidence suggesting that deductive arguments are gradually losing ground to inductive arguments as the dominant form of argumentation in academic philosophy. Similarly, Fletcher et al. (2021) find that the proportion of papers published in *Philosophical Studies* that use probabilistic methods, as opposed to formal methods, increased threefold during the first decade of the twenty-first century.

and two inductive arguments? If the two inductive arguments are only meant to support the premises of the deductive argument, should the deductive argument carry more weight than each inductive argument? These questions are beyond the scope of this paper, so we have to leave them to future studies.

5. Conclusion

In this paper, we set out to investigate the following question empirically: if science relies heavily on induction, does philosophy of science rely heavily on induction as well? Using data mining and text analysis methods, we examined a large corpus of philosophical texts mined from the JSTOR database (n = 14,199). If philosophy of science relies heavily on induction, just as science supposedly does, then we would expect to find significantly more inductive arguments than deductive arguments and abductive arguments in the published works of philosophers of science. Using indicator words to classify arguments by type (namely, deductive, inductive, and abductive arguments), we searched through our corpus to find patterns of argumentation. Overall, the results of our study suggest that philosophers of science do rely on inductive inference. But induction may not be as foundational to philosophy of science as it is thought to be for science, given that philosophers of science make significantly more deductive than inductive arguments. Interestingly, our results also suggest that philosophers of science do not rely on abductive arguments all that much, even though philosophers of science consider abduction to be a cornerstone of scientific methodology.

Acknowledgments

This paper was presented at the Digital Studies of Digital Science (DS²) online conference (UCLouvain, March 15-18, 2021). We are grateful to the organizers, Charles Pence and Luca Rivelli, and the audience for their helpful feedback. We are also very grateful to two anonymous reviewers of *Synthese* for helpful comments on earlier drafts of this paper.

References

- Ashton, Z. and Mizrahi, M. (2018). Show Me the Argument: Empirically Testing the Armchair Philosophy Picture. *Metaphilosophy* 49 (1-2): 58-70.
- Baronett, S. (2016). *Logic*. New York: Oxford University Press.
- Brigandt, I. (2014). Philosophy of Biology. In S. French and J. Saatsi (eds.), *The Bloomsbury Companion to the Philosophy of Science* (pp. 246-267). New York: Bloomsbury.
- Chakravartty, A. (2017). Scientific Realism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition).
<https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>.
- Copi, I. M., Cohen, C., and McMahon, K. (2014). *Introduction to Logic*. Fourteenth Edition. Upper Saddle River, NJ: Prentice Hall.

- Currie, G. (2016). Does Great Literature Make Us Better? In P. Catapano and S. Critchley (eds.), *The Stone Reader: Modern Philosophy in 133 Arguments* (pp. 198-202). New York: W. W. Norton & Co.
- Douglas, H. (2017). Science, Values, and Citizens. In M. P. Adams, Z. Biener, U. Feest, and J. A. Sullivan (eds.), *Eppur si muove: Doing History and Philosophy of Science with Peter Machamer* (pp. 83-96). Cham: Springer.
- Douven, I. (2017). Abduction. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition). <https://plato.stanford.edu/archives/sum2017/entries/abduction/>.
- Fletcher, S. C., Knobe, J., Wheeler, G., and Woodcock, B. A. (2021). Changing Use of Formal Methods in Philosophy: Later 2000s vs. Late 2010s. *Synthese*. <https://doi.org/10.1007/s11229-021-03433-9>.
- Govier, T. (2013). *A Practical Study of Argument*. Enhanced Seventh Edition. Belmont, CA: Wadsworth, Cengage Learning.
- Gower, B. (1997). *Scientific Method: An Historical and Philosophical Introduction*. London: Routledge.
- Henderson, L. (2020). The Problem of Induction. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition). <https://plato.stanford.edu/archives/spr2020/entries/induction-problem/>.
- Hurley, P. J. (2016). *Logic: The Essentials*. Stamford, CT: Cengage Learning.
- Hurley, P. J., and Watson, L. (2018). *A Concise Introduction to Logic*. Thirteenth Edition. Boston, MA: Cengage Learning.
- Hyland, K. (2005). *Metadiscourse*. New York: Continuum.
- Lackey, M. (2005). Review of *A House Divided: Comparing Analytic and Continental Philosophy*. *The Journal of Speculative Philosophy* 19 (3): 176-280.
- Landis, J. R., Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33 (1): 159-174.
- Lepore, E., and Cumming, S. (2013). *Meaning and Argument: An Introduction to Logic through Language*. Malden, MA: Blackwell.
- McCain, K. and Poston, T. (2017). Best Explanations: An Introduction. In K. McCain and T. Poston (eds.), *Best Explanations: New Essays on Inference to the Best Explanation* (pp. 1-4). New York: Oxford University Press.

- McMullin, E. (1992). *The Inference that Makes Science*. Milwaukee: Marquette University Press.
- Mizrahi, M. (2020). Hypothesis Testing in Scientific Practice: An Empirical Study. *International Studies in the Philosophy of Science* 33 (1): 1-21.
- Moore, G. E. (1954). *Some Main Problems of Philosophy*. London: Routledge.
- Morrow, D. R. and Weston, A. (2011). *A Workbook for Arguments: A Complete Course in Critical Thinking*. Indianapolis: Hackett Publishing Co.
- Okasha, S. (2016). *Philosophy of Science: A Very Short Introduction*. 2nd Ed. Oxford: Oxford University Press.
- Overton, J. A. (2013). "Explain" in Scientific Discourse. *Synthese* 190 (8): 1383-1405.
- Rorty, R. (2006). *Take Care of Freedom and Truth Will Take Care of Itself: Interviews with Richard Rorty*. Edited by Eduardo Mendieta. Stanford, CA: Stanford University Press.
- Russell, B. (1912). *The Problems of Philosophy*. London: William and Norgate.
- Salager-Meyer, F. (1994). Hedges and Textual Communicative Function in Medical English Written Discourse. *English for Specific Purposes* 13 (2): 149-170.
- Salmon, M. H. (2013). *Introduction to Logic and Critical Thinking*. Sixth Edition. Boston, MA: Wadsworth.
- Thabet, R. A. (2018). A Cross-Cultural Corpus Study of the Use of Hedging Markers and Dogmatism in Postgraduate Writing of Native and Non-native Speakers of English. In K. Shaalan, A. E. Hassanien, and F. Tolba (eds.), *Intelligent Natural Language Processing: Trends and Applications* (pp. 677-710). Gewerbestrasse: Springer.
- Trout, J. D. (1998). *Measuring the Intentional World*. New York: Oxford University Press.
- Wielenberg, E. J. (2015). The Parent-Child Analogy and the Limits of Skeptical Theism. *International Journal for Philosophy of Religion* 78 (3): 301-314.

Appendix I. Text Mining Methods in R

This study utilized R language and the RStudio integrated development environment for data processing. Several pre-built R packages were also used. The original corpus of JSTOR documents included a .txt file containing the full-text of the philosophical work, along with a corresponding .xml file, containing the metadata for each full-text document. It should be noted that the .xml files were converted into .txt files using the Windows Command Prompt using the command `rename *.xml *.txt` after navigating to the target folder. The command will change all .xml files in a folder to .txt.

Windows Command Prompt:

```
C:\Users\mad4\Desktop\metadata> rename *.xml *.txt
```

The *readtext()* package was used to load the full-text .txt files, as well as the converted metadata files into the RStudio environment. The *readtext* function accepts a folder path as an input parameter, i.e., *readtext*("folderpath"). The *readtext()* function will then load all files in the target folder into RStudio as a dataframe. The data frame consists of two columns: the first column is titled "doc_id" and it includes the file names from the input folder as individual elements within a string vector. The second column is titled "text" and it contains the full-text character data from each of the individual text files as a single character string. The result is a vector of character strings, with each string element containing the full-text of an individual full-text file from the input folder.

```
> library(readtext)

# Load the full text files into a string vector.

> full_text = readtext("C:/Users/mad4/Desktop/full_text_documents")

# View a summary of the data frame produced by readtext().

> glimpse(full_text)

Rows: 435,703
Columns: 2
$ doc_id <chr> "book-chapter-10.1163_j.ctt1w8h0ph.11.txt"~
$ text <chr> "<plain_text> <page sequence=\"1\"> CHAPTER 8 On
Statehood Failed Stat~
```

The converted .txt (formerly .xml) metadata files were joined to the corresponding full-text by the "doc_id" column.

```
# Load the converted metadata files into a data frame.

metadata = readtext("C:/Users/mad4/Desktop/metadata_documents")

# join the full-text and metadata records by their shared "doc_id"

corpus = left_join(full_text, metadata, by = "doc_id")
```

```

# View the corpus

glimpse(corpus)
Rows: 435,703
Columns: 2
$ doc_id <chr> "book-chapter-10.1163_j.ctt1w8h0ph.11.txt"~
$ text.x   <chr> "<plain_text> <page sequence=\"1\"> CHAPTER 8 On
Statehood Failed Stat~
$ text.y   <chr> "<book xmlns:oasis=\"http://docs.oasis-
open.org/ns/oasis-exchange/table~

```

Metadata was extracted based on XML tags found throughout the metadata records to pull information such as the journal title and publication year (i.e., <year>; </year>). The metadata was then bound to the dataframe to create a master dataframe with columns for the full-text, full metadata files, and columns for the extracted metadata.

```

# Extract the publication year.

> year_extracted = corpus$text.y %>%
  str_extract("<year>(.*?)</year>")

# Clean the XML tags from the years.

> year_clean = year_extracted %>%
  str_remove("<year>") %>%
  str_remove("</year>")

# Convert year to numeric.

> year = as.numeric(year_clean)

# Extract the journal titles and remove XML tags.

> journal_title_extracted = journal_title_test %>%
  str_remove(">") %>%
  str_remove("</journal-title")

# Convert the titles to lower-case and ensure character vectors are

```

```
not converted to factors.
```

```
> journal_title = to.lower(data.frame(journal_title_extracted,  
stringsAsFactors = FALSE))
```

Once the corpus had been assembled in RStudio, we took steps to eliminate stop words from the full-text articles. The list of stop words was generated using the `stopwords("en")` function from the “tm” package. The `stopwords()` function contains a list of 174 words commonly considered irrelevant for text-mining purposes. We chose this list due to the popularity of the “tm” package and made modifications by removing a few words from the list, including the words “so,” “for,” “of,” “the,” and “that,” since these words feature in some of our indicator pairs (see Table 2), and thus keep those words in the full-text articles. Removing the stop words from the full-text effectively shortened the ranges between words, resulting in an increase in positively matched indicator pairs in the full-text articles.

In addition to the removal of stopwords, certain stemmed words from the list of indicator pairs were replaced within the full-text. This was done to account for stemmed variations of the words which make up the indicator pairs that might also indicate the presence of an argument. For example, the argument indicator “follows” has several stemmed variations, including the past tense of the verb “follow,” which can still indicate the presence of an argument. In the case of “follows,” the words “follow,” “followed,” and “following” were replaced in the full-text with “follows” so the detection algorithm would include those words when the roots to the corresponding anchors for the indicator pairs are within the specified word-range.

We also employed this method because the indicator-pair detection algorithm searches for exact matches, and cannot be easily stemmed or lemmatized. While stemming and lemmatization functions exist in R, they could only be applied to the full-text as a whole, resulting in changes to many of the words in the full-text and not the target specified indicator words within the full-text. The result of this was an increased possibility of the detection algorithm returning false positives in cases where words with the same root but different meanings could be counted. This concern led us to take the more precise approach of building a list of words from the indicator pairs to be stemmed and ultimately replaced in the full-text. See Table 6 for a full list of the stemmed words.

Table 6. Stem-words from the indicator pairs and their corresponding replacement for the detection algorithm

Stem to be changed	Changed to match detection algorithm
absolute	absolutely
certain	certainly

definite	definitely
necessary	necessarily
prove	proves
proved	proves
proving	proves
follow	follows
followed	follows
following	follows
infers	infer
inferred	infer
inferring	infer
concludes	conclude
concluded	conclude
concluding	conclude
probable	probably
improbably	improbable
likelihood	likely
unlikelihood	unlikely
accounts	account
accounted	account
accounting	account
accountable	account
explains	explain
explained	explain
explaining	explain
explainable	explain

makes	make
made	make
making	make

The journals were then filtered by journal title and the total number of articles per journal across the entire corpus of documents was calculated. The number of articles per journal was also calculated by argumentation-type and word-range. A pattern-matching search was employed to assign the argumentation-type to documents. This was done using a regular expression as a parameter for the *str_detect()* function. The *str_detect()* function searches the full-text of each article for a specific pattern. In this case, the regular expression is used to define a complex search pattern for the presence of an indicator-pairing commonly used to identify argumentation types in academic philosophy. The pattern searches for the root-word and an anchor-word or phrase. For example, the indicator-pair of “hence account for” will search for “hence” as the root and “account for” as the anchor:

```
(?:root\\W+(?:\\w+\\W+){0,6}?anchor|anchor\\W+(?:\\w+\\W+){0,6}?root)
```

The number of words permitted, exclusively, is also specified. In the example above, the permitted number of words between the root and anchor is 6 words, indicated by the “{0,6}.” It should also be noted that the regular expression also allows the root word to precede or follow the anchor word or phrase, the order of terms does not affect detection.

```
> articles_w_matches = corpus[ 0, 2] %>% str_detect(
  regex("(?:therefore\\W+(?:\\w+\\W+){0,3}?consequently|consequently\\W
  +(?:\\w+\\W+){0,3}?therefore)"))
```

The *str_detect()* function will search each full-text character string within the full-text vector for the presence of the specified pattern and return a list of logicals where “TRUE” indicates the presence of the root-anchor pair within the specified word-range and “FALSE” suggesting the pattern is not present. These logicals are converted to character data and then to numeric, with 1 representing “TRUE” and 0 representing “FALSE.” The articles containing the positively matched patterns are then summed to yield the total number of articles within the corpus containing the pattern, as well as a separate data frame containing the full-text for each of the matched articles. These matched full-text article data frames were compiled by word-range and the argumentation-type was assigned. From these master data frames the selected journals were then filtered and summed to yield the total number of matched articles by word-range and argumentation-type for each journal.

```
> logicals_to_numeric = as.character(articles_w_matches) %>%
  str_replace("TRUE", "1") %>%
  str_replace("FALSE", "0") %>%
```

```
as.numeric()
```

This method has some limitations. First, the *str_detect()* function can only determine the presence of a specified pattern one time for each article. It is therefore possible that some articles contained the same root-anchor pattern more than once within a single article. Multiple occurrences per article were not counted. Second, the manner in which the master data frames were assembled allowed an article to repeat more than once within a list. This could happen if a single article matched different root-anchor pairs within the same word-range. For example, if article *x* contains the root-anchor pairings of “therefore absolutely” and “hence necessarily,” where both root and anchor are within the 3 words, article *x* would then appear twice within the list for matched deductive articles within the 3-word range. In this way, it is also possible that article *x* may contain more than one argumentation type and can appear in multiple lists.

To calibrate the pattern detection algorithm, we selected a small sample of 12 articles (2 articles from each journal times six journals equals 12 articles) that frequently matched the algorithm and filtered through the data to return the total number of detected arguments by argument-type and word-range for each article. Each of the expert coders read and manually identified the presence of each argument-type in each of the 12 articles in the small sample, keeping a total of the number of arguments identified by each argument-type. We then compared the findings of the coders with the search results of the algorithm.

Generally, the coders identified more arguments per article in the small sample of 12 articles than the algorithm did. In addition, the algorithm detected more inductive arguments per article in the small sample of 12 articles than the coders did. This is especially true for deductive arguments in the small sample of 12 articles. The mean ratio of deductive arguments detected by the algorithm is 0.60, whereas the mean ratios of deductive arguments detected by the three coders are 0.80, 0.83, and 0.80, respectively. Also, the mean ratio of inductive arguments detected by the algorithm is 0.34, whereas the mean ratios of inductive arguments detected by the three coders are 0.15, 0.12, and 0.14, respectively. There are likely a couple of reasons for this. First, unlike the algorithm, the coders were not limited by word range. The algorithm detects arguments by type with 3 words, 6 words, and 10 words between a root and an anchor in an indicator pair, whereas the coders can detect arguments with more than 10 words between a root and an anchor. Second, unlike the algorithm, the coders were not limited to identifying arguments by indicator words for arguments. The algorithm detects arguments based on a list of indicator pairs that contain conclusion indicators, such as ‘therefore’ and ‘conclude’ (see Table 2), whereas the coders know that “the mere occurrence [or absence] of an indicator word does not guarantee the presence [or absence] of an argument” (Hurley 2016, p. 23), and so they could identify arguments in articles even if there are no indicator words at all in the text. Finally, unlike the algorithm, the coders were not limited by indicator words for *types* of arguments. The algorithm detects arguments by *type* based on a list of indicator pairs that contain argument-type indicators, such as ‘necessarily’ and ‘probably’ (see Table 2), whereas the coders might decide to classify an argument as deductive even if it occurs in the context of inductive indicators, such as ‘probably’ and ‘likely’, or classify an inductive argument as deductive even if it occurs in the context of deductive indicators, such as ‘necessarily’ and ‘certainly’. These limitations of the pattern detection algorithm, then, may have resulted in less positive matches of deductive

arguments by the algorithm than the coders, and more positive matches of inductive arguments by the algorithm than the coders.

Nevertheless, both the coders and the pattern detection algorithm tend to find more deductive arguments than inductive arguments, and more inductive arguments than abductive arguments overall. In terms of proportions, that is, both the algorithm and the coders identified more deductive arguments than inductive arguments, and more inductive arguments than abductive arguments. Since the coders were not limited by word range in their task of identifying arguments in the small sample of 12 articles, as mentioned above, we compared their findings to the search results of the algorithm for the 10-word search results. The results of a one-way ANOVA indicated that the proportions of deductive, inductive, and abductive arguments in the small sample of 12 articles are unequal, $F(2, 33) = 60.67, p < .001$, just as they are as far as the results of the three coders are concerned: $F(2, 33) = 669.94, p < .001$ for the first coder, $F(2, 33) = 291.86, p < .001$ for the second coder, and $F(2, 33) = 293.87, p < .001$ for third coder. The proportions of deductive, inductive, and abductive arguments detected by the algorithm and the coders in the small sample of 12 articles are summarized in Table 7.

Table 7. Proportions of deductive, inductive, and abductive arguments detected by the search algorithm and the three expert coders (C1, C2, and C3) in the sample of 12 articles

Article	Alg de	Alg in	Alg ab	C1 de	C1 in	C1 ab	C2 de	C2 in	C2 ab	C3 de	C3 in	C3 ab
1	0.47	0.47	0.07	0.79	0.18	0.04	0.89	0.11	0.00	0.89	0.11	0.00
2	0.67	0.27	0.07	0.85	0.10	0.05	0.91	0.06	0.03	0.90	0.07	0.03
3	0.64	0.27	0.09	0.90	0.07	0.03	0.91	0.06	0.03	0.88	0.08	0.04
4	0.64	0.27	0.09	0.79	0.21	0.00	0.92	0.08	0.00	0.94	0.06	0.00
5	0.86	0.00	0.14	0.86	0.14	0.00	0.93	0.07	0.00	0.89	0.11	0.00
6	0.63	0.38	0.00	0.89	0.07	0.04	0.98	0.00	0.02	0.89	0.08	0.03
7	0.60	0.40	0.00	0.76	0.17	0.07	0.69	0.21	0.10	0.64	0.23	0.13
8	0.55	0.36	0.09	0.75	0.17	0.08	0.71	0.18	0.11	0.71	0.19	0.10
9	0.71	0.21	0.07	0.67	0.25	0.08	0.69	0.31	0.00	0.65	0.29	0.06
10	0.59	0.35	0.06	0.79	0.12	0.09	0.94	0.04	0.02	0.84	0.09	0.07
11	0.42	0.50	0.08	0.75	0.19	0.06	0.65	0.23	0.13	0.65	0.23	0.12
12	0.40	0.60	0.00	0.88	0.09	0.03	0.86	0.10	0.05	0.80	0.15	0.05

These results suggest that, while there are differences in the numbers of arguments by type detected by the algorithm and the coders, the general patterns are the same. That is to say, both the algorithm and the coders detected more deductive arguments in proportion to inductive arguments, and more inductive arguments in proportion to abductive arguments overall. In addition, we also checked for interrater reliability among the pattern detection algorithm and the three expert coders. See Appendix II.

Appendix II. Assessing Inter-Rater Reliability

As suggested by an anonymous reviewer, we also calculated Fleiss' kappa by comparing the results of the pattern detection algorithm when applied to the small sample of 12 articles. This was done using the counts returned by the pattern detection algorithm across the 10-word data range on the 12 articles reviewed by the expert coders and the counts from each of the expert coders' (M, A, and D) results. This creates 144 data points to compare between the four raters.

To calculate Fleiss' kappa, the data was transformed from numeric to categorical data, with any articles containing positively detected arguments changed to the argument-type (i.e., deductive, inductive, or abductive). Articles which were rated as not having an argument-type were changed to "no argument."

```
> glimpse(fleiss_categorical)
Rows: 36
Columns: 4
$ alg_10 <fct> deductive, inductive, abductive, deductive, inductive~
$ M      <fct> deductive, inductive, abductive, deductive, inductive~
$ A      <fct> deductive, inductive, no argument, deductive, inducti~
$ D      <fct> deductive, no argument, no argument, deductive, induc~
```

The `kappam.fleiss()` function from the "irr" package was used to calculate the kappa. Fleiss' kappa was calculated from among the four raters, among the three expert coders together, between the algorithm and each individual expert coder, and between each of the expert coders. The results suggest that there is substantial agreement between the algorithm and all of the expert coders (Landis and Koch 1977). Among the three coders, there is almost perfect agreement (see Table 8).

Table 8. Interrater reliability for the pattern detection algorithm and the three expert coders who have identified arguments by type in the small sample of 12 articles

Rater Comparison	Fleiss' kappa	Measure of agreement
4 Raters	0.795	Substantial agreement
3 Expert Coders	0.847	Almost perfect agreement
Algorithm - Coder M	0.803	Substantial agreement

Algorithm - Coder A	0.692	Substantial agreement
Algorithm - Coder D	0.732	Substantial agreement
Coder M - Coder A	0.883	Almost perfect agreement
Coder M - Coder D	0.846	Almost perfect agreement
Coder A - Coder D	0.81	Almost perfect agreement